DOCUMENT RESUME

ED 142 440 SE 022 909

AUTHOR Herbert, Martin
TITLE Evaluation Report 2-B-1. Second Grade Test Data.
INSTITUTION Central Midwestern Regional Educational Lab., Inc., St. Louis, Mo.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Oct 75
NOTE 77p.; Extended Pilot Trial of the Comprehensive School Mathematics Program; For related documents, see SE 022 895, 908-916 and ED 101 993-ED 102 007; Marginal legibility in Tables

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS *Curriculum; Elementary Grades; *Elementary School Mathematics; *Grade 2; Instruction; Instructional Materials; Mathematics; Mathematics Education; *Program Evaluation; *Tests
IDENTIFIERS *Comprehensive School Mathematics Program

ABSTRACT

The Comprehensive School Mathematics Program (CSMP) is a program of CEMREL, Inc., one of the national educational laboratories, and is funded by the National Institute of Education. Its major purpose is the development of curriculum materials in mathematics for grades K-6. Beginning in September, 1973, CSMP began an extended pilot trial of its Elementary Program. During the second year, 1974-75, thirteen classes continued use of CSMP in the second grade. In May and June, 1975, another series of tests was administered to the second graders and to their comparison classes. This report describes these tests and the resulting data. All differences that were significant favored CSMP. Results were consistent across school districts and ability levels. Some learning problems with the materials are identified. (RH)

EXTENDED PILOT TRIALS OF THE COMPREHENSIVE SCHOOL MATHEMATICS PROGRAM:

EVALUATION REPORT SERIES

Evaluation Report 2-B-1
Second Grade Test Data

cemrel

Extended Pilot Trial of the Comprehensive School Mathematics Program

Evaluation Report 2-B-1

SECOND GRADE TEST DATA

Martin Herbert

October, 1975

Developed by CEMREL, Inc., a private nonprofit corporation supported in part as an educational laboratory by funds from the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement should be inferred.

Description of Evaluation Report Series

The Comprehensive School Mathematics Program (CSMP) is a program of CEMREL, Inc., one of the national educational laboratories, and is funded by the National Institute of Education. Its major purpose is the development of curriculum materials for grades K-6.

Beginning in September, 1973, CSMP began an extended pilot trial of its Elementary Program. The pilot trial is longitudinal in nature; students who began using CSMP materials in kindergarten or first grade in 1973-74, were able to use them in first and second grades respectively in 1974-75, and will be able to use them in second and third grades in 1975-76. Hence the adjective "extended". The limited scope of these trials does not justify the term "field trial" since the major focus of the evaluation is on a limited number of classes in the metropolitan St. Louis area.

The evaluation of the program in this extended pilot trial is intended to be reasonably comprehensive and to supply information desired by a wide variety of audiences. For that reason the reports in this series are reasonably non-technical and do not attempt to widely explore some of the related research issues. The list of reports from the first two years of the extended pilot trial is given on the next page. The most comprehensive of these are the following:

1-A-1: Overview, Design and Instrumentation
1-A-3: Final Summary Report, Year 1
and 2-A-1: Final Summary Report, Year 2

The first of these will be particularly useful to the reader in providing a description of the program, the philosophy and goals of the evaluation and the relationship of individual reports to the evaluation effort as a whole.

Longitudinal Pilot Study of the
Comprehensive School Mathematics Program

Evaluation Report Series

Evaluation Report 1-A-1	Overview, Design and Instrumentation
Evaluation Report 1-A-2	External Review of CSMP Materials
Evaluation Report 1-A-3	Final Summary Report Year 1
Evaluation Report 1-B-1	Mid-Year Test Data: CSMP First Grade Content
Evaluation Report 1-B-2	End-of-Year Test Data: CSMP First Grade Content
Evaluation Report 1-B-3	End-of-Year Test Data: Standard First Grade Content
Evaluation Report 1-B-4	End-of-Year Test Data: CSMP Kindergarten Content
Evaluation Report 1-B-5	Test Data on Some General Cognitive Skills Related to CSMP Content
Evaluation Report 1-B-6	Summary Test Data: Detroit Schools
Evaluation Report 1-C-1	Teacher Training Report
Evaluation Report 1-C-2	Observations of CSMP First Grade Classes
Evaluation Report 1-C-3	Mid-Year Data from Teacher Questionnaires
Evaluation Report 1-C-4	End-of-Year Data from Teacher Questionnaires
Evaluation Report 1-C-5	Interviews with CSMP Kindergarten Teachers
Evaluation Report 1-C-6	Analysis of Teacher Logs
Evaluation Report 2-A-1	Final Summary Report Year 2
Evaluation Report 2-B-1	Second Grade Test Data
Evaluation Report 2-B-2	Readministration of First Grade Test Items
Evaluation Report 2-B-3	Student Interviews
Evaluation Report 2-C-1	Teacher Questionnaire Data
Evaluation Report 2-C-2	Teacher Interviews, Second Grade
Evaluation Report 2-C-3	Teacher Interviews, First Grade

Key to Indexing

1-C-2 Observations of CSMP First Grade Classes

- "2" refers simply to the number within a given year and type of data
- "C" refers to the type of data being reported
  - A: Overview, summary and theoretical reports
  - B: Student outcomes
  - C: Non-test data
- "1" refers to the year of the Pilot Study according to the following:

	Kindergarten	First Grade	Second Grade	Third Grade	. . .
Year 1 (1973-74)					
Year 2 (1974-75)					
Year 3 (1975-76)					

Table of Contents

Introduction

In the fall of 1973, the Comprehensive School Mathematics Program (CSMP) began a longitudinal pilot study of its Elementary Program. Over 100 teachers began using the program, either in first grade or kindergarten. During the 1974-75 school year just past, the second year of this pilot study, most of these classes continued into second grade and first grade respectively and many new classes began using CSMP materials.

For the purposes of the pilot study, classes in the St. Louis area are designated "local". For these classes teacher training is standardized and comparison classes established. These local classes provide much of the evaluation data derived from the pilot study including data related to classroom observations, student and teacher interviews and individualized testing. Classes not in the St. Louis area are designated "outer ring". These classes provide information concerning usage of materials (via questionnaires and teacher logs) and various corroborative test data from cooperating sites.

In the first year of the pilot study, 1973-74, a rather extensive series of tests was administered to local first grade CSMP classes and to their paired comparison classes.* During the second year of the pilot study just passed, 1974-75, thirteen of these classes continued as reasonably intact groups to study CSMP in the second grade. In May and June another series of tests was administered to these now second graders and to their comparison classes, which were, with one exception, the same groups of students used for comparison in the first year of the pilot study. This report describes these tests and the resulting data.

*Evaluation Report 1-B-2: End-of-Year Test Data: CSMP First Grade Content, Evaluation Report 1-B-3: End-of-Year Test Data: Standard First Grade Content, Evaluation Report 1-B-5: Test Data on Some General Cognitive Skills Related to CSMP Content

First Year Results and Second Year Setting

First Year Results

In the first year of the Extended Pilot Study, 1973-74, the focus was on the kindergarten and first grade classes in the local St. Louis area, particularly the first grade classes. While much data dealing with the implementation of the program was collected including teacher logs, questionnaires, classroom observations, cost figures, etc. and while much of the above, in addition to some achievement data, was collected from "outer ring" (distant from St. Louis) sites, these are summarized in full elsewhere* and will not be reviewed here. For the purpose of this report, what will be reviewed here is the design and result of testing carried out with local first grade classes. The reader who is familiar with these results may wish to proceed directly to the next section of this chapter (p. 6).

There were 16 local first grade classes located in five school districts as shown in Table 1.

Table 1

Description of Local
First Grade CSMP Classes, 1973-74

	Class Number	Mean Number of Students/Class	Predominant Racial Makeup			Estimated Socio-Economic Status					Type of Community
			Black	White	Mixed	Low	Low-Middle	Middle	High Middle	High	
District 1	1-4	32	✓		✓	✓	✓				Urban
District 2	5-8	22			✓		✓		✓		Suburban
District 3	9-10	24		✓				✓			Suburban
District 4	11-15	25		✓			✓	✓			Rural/Suburban
District 5	16	26		✓			✓	✓			Small Town

*Evaluation Report: 1-A-1: Overview, Design and Instrumentation

For each of these 16 CSMP classes, a comparison class was established, in the same school if possible or in an adjacent school if the CSMP class was the only first grade class in a school (4 cases). In the fall, the Cognitive Abilities Test (Houghton Mifflin, 1968) was administered to all 32 classes and served as a covariate for adjusting end-of-year test scores.

Three kinds of tests were administered in late spring 1974 by specially trained testers. Each of the tests is more fully described in the referenced evaluation report.

a) Tests of Standard Content (Evaluation Report 1-B-3): An analysis was made of five of the largest selling commercial math series in order to determine what skills and concepts are generally taught to first grade students. Emphasis was placed on the actual tasks provided for students, usually in the form of practice exercises. Based on this analysis, 8 test scales were constructed, each one covering a different content area and generally using the kind of item format commonly found in the text books. For each local first grade CSMP and comparison class, a random half of the students took part of the test; the other half of each class took the rest of the test.

   For each test scale, the class mean was calculated for each of the 16 CSMP and comparison classes. Table 2 shows the resulting data using an analysis of covariance procedure which adjusted scores for differences in entering ability as measured by the pretest. It can be seen that CSMP classes had significantly ($p<.05$) higher scores on the Larger Number scales and the difference approached significance on two other scales, Place Value and Measurement. The only scale on which comparison classes had higher scores was the Subtraction scale and the difference was not significant.

Table 2

Analyses of Class Means
for Tests of Standard Content

Subtest	Number of Items	Adjusted Mean Scores*		F Test** P less than
		CSMP	Non-CSMP	
Numeration	12	9.8	9.5	.22
Subtraction	12	7.4	8.4	.13
Order	8	5.6	5.4	.52
Applications	7	5.0	5.0	.97
Larger Numbers (A)	7	3.0	2.3	.01
Larger Numbers (B)	4	0.9	0.4	.01
Place Value	7	3.4	2.8	.06
Measurement	14	8.5	7.8	.07
Addition	13	10.2	9.9	.53

* Adjusted for entering ability based on pretest
** 1 and 14 degrees of freedom

b) Tests of Cognitive Skills (Evaluation Report 1-B-5): Three tests were constructed which dealt with situations novel to first grade students but which were thought to be conceptually related to some of the major ideas in the CSMP program. The tests dealt with classification skills, relational concepts, and analytic thinking respectively. Each test was designed to be administered on an individual basis, requiring 10-15 minutes each. For each test, all the students in from four to six CSMP classes and in the appropriate comparison classes were administered the test.

The analysis of class means, with adjustment for pretest scores, was carried out in the usual way. The results are given in Table 3. The differences were not significant on any of the tests, although CSMP classes had higher mean scores in each case, particularly the Relational Thinking and Analytic Reasoning tests.

Table 3

Analysis of Class Means
for Three Tests of Cognitive Skills

Test	Number of Items	Adjusted Mean Scores*		F Ratio:**
		CSMP	Non-CSMP	P less than
Classification	20	14.34	13.78	.44
Relational Thinking	15	10.92	10.19	.13
Analytic Reasoning	9	3.25	2.75	.16

* Adjusted for pretest scores
** 1 and 5, 1 and 9, and 1 and 7 degrees of freedom respectively

c) Tests of CSMP First Grade Content (1-B-2): A series of 12 subtests, comprised of 121 items, was administered to the 16 local CSMP classes. The tests were based on the CSMP Content Resume, a document prepared by the evaluation staff which outlines the development of content based on the tasks contained in the workbook series. Content on which students had already shown a high degree of proficiency in the Mid-Year Test was not retested. Because of the spiral nature of the CSMP curriculum, the test items of any given content strand exhibited a wide range of difficulty levels, ranging from very easy items on which virtually all students were expected to be at mastery level, to items testing content much further along in the strand which students had very little experience with and which only the better ones might get.

A very brief summary of the results is given in Table 4. The ratings given by the evaluation staff for each content area were subjective in nature and took into account, besides the percentages correct, the number of lessons and amount of workbook practice devoted to that content, the level of difficulty of the test items and the importance of that content area in preparing for new material. It was deemed inadvisable to administer this test to comparison classes not only because of the unique languages (arrows, strings, minicomputer) of CSMP but because, based on the analysis described above, certain content areas (multiplication, integers, probability) are not taught in traditional first grade curricula.

Table 4

Summary Data from End-of-Year
Test of CSMP First Grade Content

Subtest	Number of Items	Mean % Correct	Rating of Performance*
Multiplication	10	54	A
Integers	8	66	A
Rationals (a)	11	82	VG
Rationals (b)	5	49	A
Counting Money	5	62	A
Minicomputer	17	51	A
Relations	27	60	A
Addition	15	77	A
Subtraction	13	81	VG
Order	6	64	A
Probability	3	54	A
Venn Diagrams	1	39	

*VG-Very Good, A-Adequate, I-Inadequate, VP-Very Poor

Second Year Setting

There were many new kindergarten and first grade classes started locally in 1974-75 but the evaluation emphasis was on the second grade classes of students who would be completing their second year of the CSMP curriculum. After the first year of the Extended Pilot Study, schools made the decision of whether or not to continue using CSMP in second grade with the orignial class of first graders. One school elected not to continue and in a second the intended teacher was transferred to another school too late to arrange training for her replacement. The other 14 of the sixteen first grade classes were continued with second grade CSMP materials.

In one of the 14 classes more than half the second grade students did not have a CSMP background and certain instructional materials were late in arriving, prompting the teacher to discontinue the implementation of the program.

The design was essentially the same as for the previous year. The comparison classes established the previous year were, with two exceptions, continued virtually intact as second grade comparison classes. In those two cases, new comparison classes were established. The Kuhlmann-Anderson Test, a test of mental ability, was administered to all classes in the fall, and these scores were used as covariates for adjusted scores on the basis of differing entering ability. This test will be referred to subsequently as the "pretest", though not in the sense of a pre-post test designed to measure change. As will be seen later such adjustments were very small because of the excellent matches between the CSMP classes and their paired comparison classes. Table 5 summarizes the situation. The number of students is the number present for the pretest. Some students were absent; others entered or left the class during the school year. But this generally happened in about equal numbers between CSMP and non-CSMP classes so that for comparative purposes Table 5 reflects the rather similar class sizes and mean ability scores. In the four cases where the comparison class was not located in the same school as the CSMP school (because the CSMP class was the only second grade class in the school), the comparison class was drawn from an adjacent school.

Table 5

Local Second Grade Classes
in Extended Pilot Study, 1974-75

District Number	Class Number	CSMP Class		Comparison Class		Classes Located in Same School?
		Mean Pretest Score	Number of Students	Mean Pretest Raw Score	Number of Students	
1	1	51.5	17	49.6	17	Yes
	2	52.4	24	51.7	23	Yes
2	3 *	62.3	25	59.6	25	Yes
	4 *	42.5	21	49.7	19	Yes
	5 *	52.4	23	49.1	20	Yes
	6 *	56.5	22	55.2	25	Yes
	7	58.9	15	58.2	28	No
3	8	54.2	25	53.5	24	Yes
	9	54.3	28	49.4	25	Yes
	10	52.5	23	49.7	27	Yes
4	11	36.1	27	31.9	28	No
	12	42.3	27	45.0	26	No
	13	45.0	25	50.4	30	No
	Mean	50.8	23.2	50.2	24.4	

*Class pair 3 and class pair 4 were located in the same school. So were class pairs 5 and 6.

Testing Plan

Testing Objectives

1. To compare CSMP and Non-CSMP students' computational skills and understanding of the content generally considered to be standard in second grade mathematics. For this purpose, a standardized test, the Comprehensive Test of Basic Skills, Level C, (California Test Bureau, 1973), was selected and administered to all second grade classes

2. To determine whether or not CSMP students were better able to solve a wide variety of mathematical "problems", not directly similar to the content of CSMP, but related in a transfer or application sense. A series of tasks was constructed, some of which were administered on an individual student basis and some to the whole class. These Comparison Tasks were grouped into four tests: Individual Test 1, Individual Test 2, Group Test 1 and Group Test 2. These were administered to various CSMP and comparison classes according to the plan given below.

3. To investigate student achievement in the unique aspects of CSMP. A series of five tests was constructed dealing with specific CSMP content, such as the Minicomputer and arrow diagrams. Areas of content already tested in 1, above, or indirectly in 2, above, were omitted. These CSMP Tests were administered to CSMP classes only, according to the plan below.

   These three groups of tests and the results for them are described in succeeding chapters.

Testing Plan

Because of classroom time demands it was not possible to administer all tests to all classes. The tests were allocated as described in Table 6, below. Each test was administered to at least six pairs of classes (CSMP - Non-CSMP) and the sample was drawn to ensure representative classes in terms of student ability and geographic area.

For the standardized and comparison tests, the X in Table 6 indicates that the test was given to both the CSMP class and its comparison class. In the case of the individually administered tests a sample of seven students was chosen from each class by the following method. Students were rank ordered on pretest scores and selected in such a way as to provide equal intervals of students between those selected. For example in a class of 28 students this might mean the students with rank 2, 6, 10, 14, 18, 22 and 26 would be selected for Individual Test 1. Not all selected students took the test because of absentees and transfers and alternates were designated when feasible. A different group of seven students was then selected for Individual Test 2.

Three testers were specially trained to administer the tests. One handled all the individual testing, the other two did the group testing. The training required about two to three days made up of several short sessions and the tests were administered during the last seven weeks of the school year. Four class periods of from 25-45 minutes were required for each class (three for Non-CSMP classes since they were not given the CSMP tests). The individual tests generally required about 20 minutes per student.

Table 6

Testing Plan for End-of-Year Tests

District	Class Pair	Standardized Test		Comparison Tests				CSMP Tests				
		Computation	Concepts and Applications	Group Test 1	Group Test 2	Individual Test 1	Individual Test 2	A	B	C	D	E
1	1	X	X		X	X	X	X			X	X
	2	X	X	X		X	X		X	X		
2	3	X	X		X	X	X			X	X	
	4	X	X	X		X	X		X		X	X
	5	X	X	X		X		X			X	
	6	X	X		X	X	X		X	X		X
	7	X	X	X		X	X	X		X		X
3	8	X	X		X	X	X	X	X			X
	9	X	X		X	X	X		X		X	X
	10	X	X	X		X	X			X	X	
4	11	X	X	X						X	X	
	12	X	X		X	X	X	X		X		X
	13	X	X		X	X	X		X	X		X
5*	14	X	X					X	X			X
	15	X	X					X			X	
	16	X	X					X	X		X	

*The three classes in this district were the only three CSMP classes and it was not possible to establish adequate comparison classes. Hence they are not technically part of the design (comparison) although scores will be reported separately for them where appropriate later in the report.

Method of Analysis

Two basic procedures were used. When making statistical comparisons between CSMP and Non-CSMP classes the class means were treated as the unit of analysis and a two-way Analysis of Covariance procedure was used with pretest scores being the covariate so that scores were adjusted for differences in entering ability. One classification had two levels; CSMP and non-CSMP. The other classification had as many levels as there were class pairings and was essentially a school classification.

Under this model one assumes that the variation in scores can be explained by the covariate (difference in entering ability), class pairing (differences in school effects), treatment (differences in math curriculum) plus unexplained error. The ratio of the variation from treatment differences to the variation from error is then the basis of an F test. One hypothesizes that there are no resulting differences due to treatment (i.e. between CSMP and Non-CSMP classes) and that resulting differences which actually did occur, occurred by chance. One can then determine from the F statistic the p-value or probability of this chance occurrence. If the probability is low (.05 is often taken as a standard), between CSMP and Non-CSMP classes and say that there are "significant" differences. It is this p-value which will be reported for each test.

For this analysis it is necessary that only students who were present for the pretest be included in the analysis. Thus, for any particular class, the pretest score for a given test will vary slightly depending on which students were absent for that test. Later in this report the performance of new-to-CSMP students, those who transferred into a CSMP class in second grade, will be compared with students who were original CSMP first graders.

Classes designated as Non-CSMP classes were simply that - not using the CSMP curriculum. These classes used one of three popular elementary mathematics texts published by Houghton Mifflin, Addison Wesley and Silver Burdett, but for the purpose of this study they have been lumped together as comparison classes as if they were studying the same curriculum. In fact these texts are very similar (compared to CSMP) in relative emphasis given to various topics (see Evaluation Report 1-B-3) and one would expect that among classes using these curricula there would be only small differences in achievement which could be attributed to a particular text.

For each test an item analysis was produced. Students across all CSMP classes were grouped together and various item and test statistics were derived. The same procedure was followed for Non-CSMP students. This enables one to compare the two groups on individual test items and see which items or groups of items were answered more successfully by one group or the other. What are provided in the item analyses section are thus merely descriptive data; no tests of significance are made. The statistical hypotheses regarding CSMP versus Non-CSMP performance are tested through analysis of class means (described in the previous section). Thus the item analysis supplements and may help in the interpretation of those results. Often in this report the item analysis precedes the analysis of class means. The reason for this is that the item analysis sections provide the actual test items and it seems useful for the reader to be able to look at the test items and determine what the test is all about before going to the statistical comparison of class means.

Standardized Tests

Description of Test

The Comprehensive Test of Basic Skills is a series of 10 standardized achievement tests. Level C is the grade range 1.6 to 2.9 and was standardized in 1973. The test is published by the California Test Bureau/McGraw Hill.

From the Administrator's Manual, Part 1, Description of CTBS:

"RATIONALE

A carefully formulated rationale formed the basis for each step in the development of all levels of CTBS, *Expanded Edition*. This rationale required that the tests measure systematically those skills prerequisite to studying and learning in school. CTBS, *Expanded Edition* is not intended to measure achievement in specific course content as reflected in textbooks for various grade levels. The tests are intended for national use, by students who have been taught according to various approaches. Test items should be answered as readily by students taught according to a traditional approach as by those who are taught according to any of the newer approaches. However, performance on these tests necessarily depends on the possession of relevant knowledge and is affected by the grade level at which a skill is first introduced. It is assumed that all curricula are formulated to increase, through the grades, a student's competence in dealing with content of increasing difficulty. These tests aim to measure, therefore, those skills common to all curricula."

and later:

"Test 7 - Mathematics Computation

The 28 items in Test 7 consist of 10 addition, 10 subtraction, and 8 multiplication problems. Each operation is tested in a separately timed section so that a measure of ability for even the slow students is provided for all operations. However, there are no separate norms; a total score for the three sections will be reported. Addition and subtraction problems are displayed in both horizontal and vertical formats. They include operations on one-, two-, and three-digit numerals. Multiplication problems are in the horizontal format, and all are single-digit multiplications.

Test 8 - Mathematics Concepts and Applications

The 25 items in Test 8 measure skills in basic operations: numbers, numeration, measurement, and fractions. The problems are read aloud to the students, who select their answers from pictured, numerical, or printed responses. One total score is reported for this test."

The Computation Test requires exactly 34 minutes and consists of 28 items and the Concepts and Applications Test requires approximately 25 minutes and consists of 25 items.

Analysis of Class Means

Table 7 gives the class means for each of the three operations and for the total of the Computation Test, for the total of the Concepts and Applications Test and for the Total Mathematics Score. Also given is the mean class pretest scores, which may vary somewhat for a given class because of student absentees (the tests were administered at different times).

Table 7

Class Mean Scores
Comprehensive Test of Basic Skills

Class Pair	Computation Test					Concepts and Applications		Total Score	Pretest
	Add^n	Sub^n	Mult^n	Total	Pretest	C. and A.	Pretest		
1	7.88 6.87	6.44 5.53	5.00 5.53	19.32 17.93	50.9 51.1	18.13 17.60	50.9 51.1	37.45 35.53	50.9 51.1
2	8.09 7.74	6.41 6.22	6.45 6.91	20.95 20.87	52.4 51.7	19.04 18.96	52.4 51.7	39.95 39.83	52.4 51.7
3	8.76 7.76	9.71 8.72	7.33 7.48	25.80 23.96	61.1 60.5	23.14 22.46	61.1 61.1	48.95 46.66	61.1 61.1
4	6.40 6.56	6.13 5.38	6.07 5.50	18.60 17.44	42.4 50.8	16.28 17.38	43.5 50.8	34.73 34.82	42.4 50.8
5	8.26 7.07	8.00 7.88	6.05 7.13	22.31 22.07	52.1 52.3	19.26 19.94	52.1 51.8	41.57 41.88	52.1 52.3
6	9.00 8.49	8.71 8.81	6.94 6.24	24.65 23.53	55.7 57.4	21.00 20.36	55.7 56.9	45.65 44.34	55.7 57.4
7	8.23 7.30	8.77 7.37	7.46 6.89	24.46 21.56	60.2 58.8	20.46 18.85	60.2 58.8	44.92 40.41	60.2 58.8
8	8.21 7.25	8.25 8.40	6.67 6.00	23.13 21.65	54.4 54.1	21.45 20.25	55.1 54.3	44.90 42.21	55.1 54.6
9	7.92 6.72	8.38 6.59	7.38 6.77	23.68 20.08	54.2 49.0	21.21 16.86	53.6 48.6	44.70 36.81	52.8 48.6
10	7.26 6.40	6.70 7.00	6.00 5.92	19.76 19.32	52.5 50.0	17.83 15.60	52.5 50.0	37.59 34.92	52.5 50.0
11	6.00 5.42	5.33 4.33	5.06 3.75	16.39 13.50	36.9 31.9	15.78 10.91	36.9 32.2	32.17 24.52	36.9 32.2
12	6.92 6.45	7.29 6.64	6.21 6.05	20.42 19.14	42.7 44.4	16.65 17.82	42.6 44.4	37.43 36.96	42.6 44.4
13	7.57 7.68	7.05 8.14	7.05 6.82	21.67 22.64	47.0 50.4	16.29 18.61	47.0 50.4	37.96 41.25	47.0 50.4

Table 8, below, gives the mean score across the 13 CSMP classes and across the 13 Non-CSMP classes for the various tests of the Comprehensive Tests of Basic Skills. Also given for the two groups are the mean pretest scores which are almost equal. The adjustment in scores due to differences in entering ability never exceeded .05 (though the pretest was effective in reducing the unexplained error variance; from 30% to 75% on the two tests and total score). Other traditional Analysis of Covariance statistics are given in the Appendix for this and subsequent series of tests. Because of the excellent match of classes the adjustments in scores due to differences in entering ability pretest scores were small, never exceeding .05 raw score units.

Table 8

Mean Scores and Significance Tests
CTBS: CSMP versus Non-CSMP

Test	Mean Test Score Across Classes		Mean Pretest Score Across Classes		F-Test, 1 and 11 degrees of freedom p less than:
	CSMP	Non-CSMP	CSMP	Non-CSMP	
Computation Test	21.63	20.28			.01
Addition	7.8	7.1	50.95	50.95	.01
Subtraction	7.5	7.0			.05
Multiplication	6.4	6.2			.28
Concepts and Applications Test	18.96	18.12	51.03	50.92	.03
Total Mathematics Score	40.61	38.47	50.88	51.03	.01

The p value given in the right hand column of Table 8 is based on the assumption that there are no real differences between CSMP and non-CSMP classes and that the actual observed differences occurred by "chance". The p value is the probability of this chance event. It can be seen that CSMP classes scored significantly higher than Non-CSMP classes on both the Computations Test and the Concepts and Applications Test, as well as on the Total Mathematics score. On the three parts of the Computations Test the differences were significant for the addition and for the subtraction items, but not for the multiplications items.

Item Analysis

Table 9, below, gives the percent correct for each item in the Computation for CSMP students and for Non-CSMP students. Unlike the previous section, data are averaged across students, not across classes, and therefore the data do not agree exactly with Table 8. Table 10 provides the same information for the Concepts and Applications Test. For brevity the optional answers to these multiple choice tests are not usually given and in Table 10, the oral directions accompanying each item have been shortened somewhat.

Table 9

Percent Correct
Computation Test

Addition Items			Subtraction Items			Multiplication Items		
Item	CSMP Students	Non-CSMP Students	Item	CSMP Students	Non-CSMP Students	Item	CSMP Students	Non-CSMP Students
1. 3 8 0 + 6	88	88	11. 38 - 6	88	82	21. 5x2=	87	88
2. 19 + 6	79	81	12. 545 - 33	74	72	22. 1x4=	86	84
3. 154 +128	74	64	13. 16 - 8	75	73	23. 2x3=	86	86
4. 27 +18	63	55	14. 768 -427	79	70			
5. $1.14 + .63	50	39	15. 55 -52	70	69	24. 3x5=	83	86
6. 13 6 52 + 27	60	49	16. 64 - 9	68	61	25. 3x4=	83	81
			17. 65-42=	83	81	26. 5x9=	72	70
7. 40+30=	91	86	18. 24-8=	76	71			
8. 7+3+2=	94	93	19. 13-8=	69	70	27. 4x4=	72	69
9. 11+7=	93	90						
10. 26+21=	79	73	20. 78-43=	68	63	28. 8x5=	78	70

	CSMP	Non-CSMP
Number of Students	258	283
Mean Score	21.70	20.65
Standard Deviation	5.89	5.63
Mean Pretest Score	50.02	50.37
Correlation with Pretest	.65	.57

There are no single items on which there are large differences between the two groups of students; in fact the largest difference in percent correct is only 11 (items 5,6). However, the differences were consistently in favor of CSMP students. On only four items did the Non-CSMP students have a higher percent correct (and then by at most 3%), even though they had a slightly higher pretest score.

Table 10

Percent Correct
Concepts and Applications Test

Item	CSMP Students	Non-CSMP Students
1. A way of showing 13: 11+2, 12+2, 11+1, 10+2	85	83
2. In which box does number word show how many boats there are: (4 picture-word pairs shown)	78	75
3. Two hundreds, three tens and six ones: 236, 263, 326, 362	78	81
4. Which shows counting by fives starting at 25?	86	75
5. Which shows counting by twos starting at 77?	69	62
6. Sold by the quart: pictures of bread, eggs, milk, oranges	73	76
7. Amount of money shows: (picture of quarter, dime, penny)	67	74
8. Same amount as a dime: (4 combinations of coins shown)	91	90
9. 19¢ purchase, change from a quarter (as above)	64	53
10. Monday is second day of week, which day is Thursday?	46	38
11. Which clock shows 15 minutes after 7?	73	78
12. Which clock shows ten thirty-five?	61	56
13. Find the numeral that is one-half of 6.	85	81
14. Which box shows three-fourths of rectangle dark?	94	92
15. Find the numeral that shows one-half of 4.	87	74
16. True number sentence: 5+7=11 8+6=14 6+5=9 3+9=13	87	87
17. True number sentence: 8=3+4 10=3+7 9=5+6 12=6+5	81	74
18. Mike scored 4, his brother 2. How to find total: 6+2 4-2 4+2 2+2	76	80
19. Frank caught 6 fish, then no fish. Which picture shows how many he caught altogether?	87	85
20. Six boys, 8 girls, how many altogether?	87	82
21. Three crayons, 6 crayons, 7 crayons. How many altogether?	85	84
22. Twenty-eight children, 5 absent. Which shows how to find number at school? 28-5, 28-3, 28+5, 23=5	72	59
23. Which number line shows 45+3?	71	60
24. Leroy gets up at seven o'clock. He gets to school two hours later. What time does he get to school?	62	63
25. Where is 8 the missing number? 15-□=7, 14-□=5, 13-□=6, 16-□=7	53	59
Number of Students	259	282
Mean Test Score	18.99	18.23
Standard Deviation	5.06	4.80
Mean Pretest Score	50.25	50.14
Correlation with Pretest	.65	.55

CSMP students did relatively best (at least more than 10% correct) on items 4, 9, 15, 22 and 23; five rather unrelated items. On only six of the 25 items, items 3, 6, 7, 11, 18 and 25, did Non-CSMP students do better (by at most 7%) and these items were also rather unrelated.

Comparison Tests

Four comparison tests were developed, two to be individually-administered and two group-administered. Each had from two to five subtests; a total of 12 subtests in all. The tests were intended to measure students' abilities to do certain tasks thought to be related to CSMP; they did not cover specific CSMP content. A wide variety of sources was investigated to determine the availability of potentially useful tests. This included published achievement tests and tests of mental ability, Kit of Reference Tests for Cognitive Factors (Educational Testing Service, 1963), problems from the California Mathematics Improvement Program, and various research journals in mathematics education. Some of the tests finally constructed were developed by the CSMP Evaluation staff, others were adapted from available instruments.

Each of the group-administered tests was tried out in four second grade classes and each of the individually-administered tests was tried with at least 12 students, never more than three from a given class. None of these classes or students was in any way involved in the pilot study. Several promising tests had to be discarded for one reason or another.

The 12 subtests listed below are essentially independent tasks, though some are related in various ways. Thus, although total scores for each of the four comparisons are also analyzed in this chapter, they are not homogeneous in testing a series of related and similar tasks and it is the subtest scores which provide the most easily interpretable information. The subtests named below will be described in detail in the next four sections (one for each test) of this chapter.

Group Administered		Individually Administered	
Ia): Word Problems	IIa): Equations-Construction	IIIa): Classification	IVa): Number Patterns
Ib): Number Puzzles	IIb): Equations-Fluency	IIIb): Binary Relations	IVb): Functions
Ic): Estimation-Calculations	IIc): Combinatorics		IVc): Probability
Id): Estimation-Largest Number	IId): Regrouping		
Ie): Showing Fractions			

Comparison Test I (Group Administered)

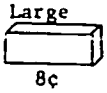
For each of the five subtests, a brief description of the tests is followed by an item analysis in which the reader can see the actual test items and the percent correct for CSMP and Non-CSMP students. Actually the items have sometimes been abbreviated, to keep reasonably concise the tables which follow. These item analyses were calculated by summing across students: 115 CSMP students who had a mean pretest score of 49.0 and 120 Non-CSMP students who had a mean pretest score of 49.5.

After the item analyses have been presented, and the reader is perhaps in a better position to understand the nature of the subtests, the Analysis of Class Means is given and this provides a statistical comparison between the six pairs of CSMP versus Non-CSMP classes.

Test Ia) Word Problems: A series of rather difficult items (mean score less than 5 out of 14 items) was read to the students, one item at a time. The student copy of the tests gave necessary pictorial information and repeated in short phrases the important part of the question. The response was open-ended; students produced the answer they thought correct. The questions dealt with what are usually called word problems (verbal problems requiring an arithmetic operation) as well as problems related to ideas of negative integers, even-odd, division, remainder, average and combining operations. Of course words like these were not used. The problem involving negative integers, for example, dealt with temperatures above and below zero.

Table 11

Item Analysis
Comparison Test Ia): Word Problems

Test Items	Percent Correct	
	CSMP	Non-CSMP
Small  3¢ Medium  5¢ Large  8¢		
1. Bill had 13¢ and bought 1 large. How much change?	70	65
2. How much do 3 smalls and 1 medium cost?	52	46
3. Mary had 20¢ and bought 2 smalls and 1 medium. How much change?	26	23
4. One large costs 1/2 as much as a sucker. A sucker costs?	11	5
5. One medium costs 4¢ less than a banana. A banana costs?	23	12
6. Four glasses fill one quart. How many quarts could be filled with 13 glasses?	41	36
7. (From above) How many glasses left?	44	46
8. A car circles a racetrack 3 times in 10 minutes. How many times in 30 minutes?	17	19
9. Each of the four snakes (picture of 4 snakes of different lengths) grows 1 inch per year. Which doubles his length first?	7	8
10. The temperature was 4 degrees below zero. Then it got 10 degrees warmer. What was the temperature then?	17	6
11. A boy is first in line, then a girl, then a boy, and so on. Is the 37th person in line a boy or a girl?	49	54
12. (From above) Is the 54th person a boy or a girl?	46	52
13. A rabbit weighed 4 ounces, then gained 2 ounces, then doubled its weight. How much did it weigh then?	26	22
14. Tom's score was 10, Bill's was 4 and John's was exactly half way between them. What was John's score?	29	20
Mean Score	4.58	4.12
Standard Deviation	3.06	2.62
Correlation with Pretest	.60	.52
KR20	.77	.70

The reader is reminded that these statistics are derived across students. Mean scores will differ by small amounts from those derived across classes in testing the significance of CSMP - Non-CSMP differences (p. 21). KR20 (bottom line) is a measure of the homogeneity of this set of test items; the degree to which they test the same underlying ability. It usually varies from 0 (zero average correlation between items) and 1 (high correlations).

CSMP students did slightly better on items 4 and 5 dealing with what might be called converse relationships, on item 10 which dealt with negative integers and on item 14, implicit averaging. These items, while they are applications of the CSMP content, are quite different from anything in the CSMP curriculum. Non-CSMP students did slightly better on items 11 and 12 regarding odd and even, though all percentages were generally close to that expected from guessing alone.

Test Ib) Number Puzzles: Given a completed addition or subtraction problem, but with one or two digits from the original question missing, students were to figure out the missing digits. For example $\begin{array}{r} 36 \\ -1\square \\ \hline 19 \end{array}$

Table 12

Item Analysis
Comparison Test Ib): Number Puzzles

Test Item	Percent Correct	
	CSMP	Non-CSMP
1. $\begin{array}{r} 5\square \\ +\square \\ \hline 17 \end{array}$	62	66
2. $2\square$	80	73
3. $+\square 6$	73	73
$\hline 58$		
4. $\square 8$	9	5
5. $+3\square$	22	11
$\hline 51$		
6. $\begin{array}{r} 36 \\ -1\square \\ \hline 19 \end{array}$	4	4
7. $5\square$	6	10
8. $-\square 7$	6	7
$\hline 24$		
Mean Score	2.62	2.50
Standard Deviation	1.33	1.22
Correlation with Pretest	.45	.50
KR20	.41	.40

Items 4-8, involving addition with carrying or subtraction with borrowing, were very difficult for students. The largest difference was on items 4 and 5, involving carrying.

Test Ic) Estimation, Calculations: On a very large sketch pad at the front of the room a calculation was shown to the class for about 10 seconds, including time for the tester to read the calculation. Students then chose the one of three alternatives shown in their test booklet which they thought closest to the answer to the calculation. For example with a calculation of ½x78 and alternatives of 30, 35, 40, one might think "78 is close to 80 and ½x80 is 40" or "2x35 is 70 and 2x40 is 80, so 40 is closer." There was not sufficient time for students to actually calculate "½x78=39 and 40 is closer to 39 than is 30 or 35." Two practice items were given before the actual test items.

Table 13

Item Analysis
Comparison Test Ic): Estimation - Calculations

Calculation Posed	Alternatives	Percent Correct	
		CSMP	Non-CSMP
1. 19+17	35, 40, 45	70	62
2. 95-34	50, 60, 70	48	43
3. 1/2x78	30, 35, 40	50	51
4. 2x63	120, 125, 135	40	38
5. 33+46	70, 80, 90	34	37
6. 96-78	20, 30, 40	26	35
7. 325+582	800, 850, 900	49	54
8. 1/2x78	30, 35, 40	45	39
9. 22+63	80, 85, 90	59	54
10. 4x19	40, 60, 80	35	45
11. 399-201	100, 150, 200	32	33
12. 1/3x63	20, 25, 30	23	13
Mean Score		5.10	5.08
Standard Deviation		2.23	1.88
Correlation with Pretest		.41	.26
KR20		.49	.25

This test was certainly a speeded test and was intended to be. Students were to "guess" about what the answer would be. Thus there was undoubtedly considerably real guessing taking place (a stab in the dark) and this is corroborated by the moderate correlation with pretest and the moderate KR20. Though there are differences between the two groups of students on individual items, the differences are not systematic according to either arithmetic operation or size of number.

Test Id) Estimation: Largest Numbers. As before, students were shown briefly two calculations, one on their left in a circle and one on their right in a square. They then had to mark on their own page the (empty) circle or square, depending on which calculation shown at the front of the room that they thought produced the largest answer.

Table 14

Item Analysis
Comparison Test Id): Estimation - Largest Number

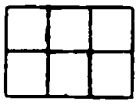
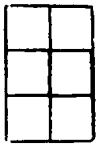
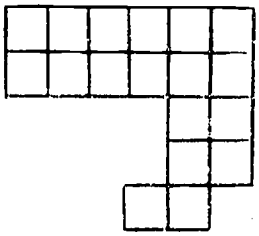
Calculations Posed (Students to select larger)	Percent Correct	
	CSMP	Non-CSMP
1. 6329 , 6328	50	54
2. 527+84 , 527+86	90	79
3. $\frac{1}{2}$x250 , $\frac{1}{3}$x250	63	21
4. 19x3 , 18x3	86	87
5. 27+80 , 26+82	62	59
6. 300-24 , 300-23	41	30
7. 19+7+5 , 5+19+8	83	79
8. 700-20 , 710-50	45	39
Mean Score	5.21	4.48
Standard Deviation	1.52	1.46
Correlation with Pretest	.48	.32
KR20	.38	.31

This subtest appears similar in some ways to the previous one. Actually it does not require computational facility, but rather what might be intuitive ideas regarding commutativity and order properties. CSMP students did better on several items, notably items 2 and 6 and especially item 3 dealing with multiplication by 1/2 or 1/3. Note that on pure calculations with 1/2 and 1/3 (items 3, 8 and 12 of subtest Ic)) there was very little difference between the two groups. This test also was characterized by moderate pretest correlation and low KR20, perhaps due in part to guessing.

Test Ie) Showing ½: For each of four shapes, usually partial grids, students were asked to shade a given fraction of the shape. For example, shade 1/3 of  .

Table 15

Item Analysis
Comparison Test Ie): Showing Fractions

Test Items (Student to shade indicated fraction of given figure)	Percent Correct	
	CSMP	Non-CSMP
1. 1/3 	10	5
2. 2/3 	40	55
3. 1/6 	21	14
4. 1/2 	43	19
Mean Score	1.13	0.9
Standard Deviation	1.00	.76
Correlation with Pretest	.45	.24
KR20	.35	.02

Item 4 required the students either to shade every other square or to count the total number of squares and shade half that many. This was the only item from the five subtests of Test I that was similar to an exercise appearing in the CSMP curriculum and there is a fairly large difference in performance for that item. Once again the correlations with pretest are moderate and the KR20, particularly for Non-CSMP students, was low.

Analysis of Class Means: Table 16 presents the important summary data in the comparison of CSMP and Non-CSMP classes for Test I. It gives for each group the mean score across the six CSMP and the six matched Non-CSMP classes for each of the subtests, for the total, and for the pretest. These means are slightly different from those in the item analyses because they were devised across classes rather than across students. Also given in Table 16 are the probability levels at which the observed differences are significant.

Table 16

Analysis of Covariance
Comparison Test I

Test	Mean Test Score Across Classes		Mean Pretest Score Across Classes		F-test, 1 and 4 degrees of freedom p less than:
	CSMP	N-CSMP	CSMP	N-CSMP	
Ia) Word Problems	4.79	4.06			.23
Ib) Number Puzzles	2.66	2.51			.56
Ic) Estimation: Calculation	5.18	5.08	49.49	49.01	.95
Id) Estimation: Largest Number	5.26	4.44			.04
Ie) Showing Fractions	1.15	0.92			.18
Total, Test I	19.06	17.01			.12

All differences were in favor of CSMP, though significance was reached at the .05 level on only subtest Ie) Estimation: Largest Number. The adjustments in mean scores due to the differences in entering ability (not shown) were small, the largest such admustment being .10 on the total score.

Comparison Test II (Group Administered)

The format of presentation is the same - first the test descriptions and item analysis for each of the subtests, then the summary comparison of class means. The former are based on 145 CSMP students with a mean pretest score of 53.14 and 146 Non-CSMP students with a mean pretest score of 52.77. Seven pairs of classes took this test.

Test IIa): Equations - Construction. Students were given a string of symbols (numerals, operation signs and equality sign) and asked to form an equation (number sentence) with them. Each symbol had to be used exactly once. For example, with the string =, +, 2, 5, 3 one could make 3+2=5 (or 2+3=5 or 5=2+3 or 5=3+2). Thus these were essentially problems with closure; to find the equations which worked.

Table 17

Item Analysis
Test IIa): Equations-Construction

Test Items (Given strings of symbols) from which equation was to be constructed)	Percent Correct	
	CSMP	Non-CSMP
1. = + 8 12 4	96	96
2. = x 6 18 3	90	91
3. = x 15 5 $\frac{1}{3}$	71	03
4. = + - 5 21 6 10	14	10
5. = - x 2 32 10 3	3	1
Mean Score	2.73	2.02
Standard Deviation	.88	.58
Correlation with Pretest	.53	.31
KR20	.46	.28

Only on item 3 was there much difference. On that item, which dealt with multiplying by $\frac{1}{3}$, more than two-thirds of the CSMP students were correct while virtually none of the Non-CSMP students were successful.

Test IIb): Equations - Fluency. This test was similar to Test IIa) in that students were to produce equations from strings of symbols. This time, however, the objective was to produce as many equations as possible using any or all of the given symbols as many times as desired. The numbers were small, the emphasis on putting combinations together appropriately rather than on computational speed. The score for each string of symbols was the number of correct equations in the 2½ minute time allowance.

Table 18

Item Analysis
Test IIb) Equations-Fluency

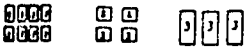
Given String of Symbols	Mean Number of Correct Equations	
	CSMP	Non-CSMP
1. = + - 2 4 6	3.27	3.55
2. = + - x 1 2 3	4.90	4.34
3. = + x $\frac{1}{2}$ 1 2 4	5.42	3.42
Total	13.57	11.31

CSMP students did better on the second item and much better on the third item, where "$\frac{1}{2}$" and "x" were given. A separate count was made in the third item of correct equations containing "x". For CSMP students the mean number was 3.33 and for Non-CSMP students it was 1.61. The mean number of correct equations using "$\frac{1}{2}$" was 1.48 for CSMP students and 0.10 for Non-CSMP students. Also on the third item, the mean number of incorrect equations produced was 1.12 for CSMP students and 2.14 for Non-CSMP students.

Test IIc): Combinatorics. Both situations given to the students asked to list the ways of doing something.

Table 19

Item Analysis
Test IIc) Combinatorics

Situations Posed	Mean Number of Ways Shown	
	CSMP	Non-CSMP
1. Given packages of gum, as pictured below (the number on the package tells how many pieces are in the package), show ways of getting exactly seven pieces. Students were actually given multiple copies of picture below and put X's on appropriate packages. 	5.01	4.99
2. List as many two digit numbers as you can from the digits 3 and 7.	11.97	12.17
Total	16.98	17.16

CSMP students might have been expected to do better on these items since they have occasional opportunities to deal with combinatoric situations. However, there was virtually no difference in scores.

Test IId): Regrouping. Two kinds of situations were given involving the filling with water of three sizes of cans: large, medium (2 mediums = 1 large), and small (2 smalls = 1 medium). In one situation, a certain number of particular-sized cans was available to fill as many of a particular larger sized can as possible, i.e., how many could be filled, and how many of the smaller cans would be left over. In the other situation given numbers and combinations of medium and large cans were shown and the student was to determine how many small cans could be filled from them. These situations also required extensive preparation by way of examples in preparing students for the situations and ways of responding. The items presented pictorially, but are given verbally in Table 20.

Table 20

Item Analysis
Test IId): Regrouping

Test Items	Percent Correct	
	CSMP	Non-CSMP
1. Given 4 smalls. How many mediums can be filled?	78	79
2. How many smalls left over?	75	71
3. Given 7 mediums. How many large can be filled?	63	71
4. How many mediums left over?	71	73
5. Given 10 smalls. How many large can be filled?	34	42
6. How many smalls left over?	38	41
7. Given 15 smalls. How many large can be filled?	23	28
8. How many smalls left over?	20	26
9. Given 17 smalls. How many mediums can be filled?	18	27
10. How many smalls left over?	35	41
11. Given 3 mediums. How many smalls can be filled?	77	78
12. Given 1 large. How many smalls can be filled?	54	52
13. Given 1 large, 2 mediums. How many smalls can be filled?	45	45
14. Given 2 large, 3 mediums. How many smalls can be filled?	32	29
15. Given 5 large, 4 mediums. How many smalls can be filled?	12	17
Total	6.76	7.20

This subtest dealt with an instance of regrouping but was perhaps only tenuously related to concepts of place value (regrouping in tens) Non-CSMP students did better on the first ten items (small units combined into larger ones) while there was no difference on the last five items (larger units broken down into smaller ones). Though the differences are not significant in either case, one might speculate that while CSMP uses primarily one pedagogical technique for place value (the Minicomputer), Non-CSMP classes may use a variety of techniques (bundles of sticks, blocks, etc.) which might or might not make them as proficient with place value in tens but which might help them more in other regrouping concepts. One of the Tests of Standard Content administered to these students in first grade dealt with filling bags with blocks, ten at a time. CSMP students did better on that task.

Analysis of Class Means: The data in Table 21 are based on seven pairs of classes.

Table 21

Analysis of Covariance
Comparison Test II

Test	Mean Test Score Across Classes		Mean Pretest Score Across Classes		F-test, 1 and 5 degrees of freedom p less than
	CSMP	N-CSMP	CSMP	N-CSMP	
IIa) Equations-Construction	2.72	2.03			.01
IIb) Equations-Fluency	13.63	11.19			.02
IIc) Combinatorics	16.94	17.08	52.46	52.15	.73
IId) Regrouping	6.34	6.55			.40
Total	39.62	36.86			.04

Significant differences were recorded on Subtests IIa) and IIb) dealing with the construction of equations and on the total of Test II. Again the adjustments in mean scores due to differences in pretest scores were small, the largest being .14 for the total score (i.e. CSMP adjusted mean = 39.48, Non-CSMP = 37.00).

Comparison Test III (Individually Administered)

The item statistics for the two subtests of Comparison Test III are based on 81 CSMP students with a mean pretest score of 53.4 and 78 Non-CSMP students with a mean pretest score of 54.1.

Test IIIa) Classification. There were two parts to this subtest. On the first part the student worked with a 4x4 array, shown below in reduced size. The figures were actually colored red or blue as indicated and were identical except for differences on the four attributes: color (red, blue), height (tall, short), width (fat, thin) and sex (boy, girl). On the second part students were given a set of 27 cards in three colors (red, blue, yellow), three shapes (square, triangle, circle) and with varying numbers of dots marked on them (one, two or three). That is they were given cards with each possible combination of attributes except that one card was missing and they were to figure out which card was missing. As before the items in the item analyses are abbreviated and do not reflect the careful explanations, repetitions and encouragement given by the tester.

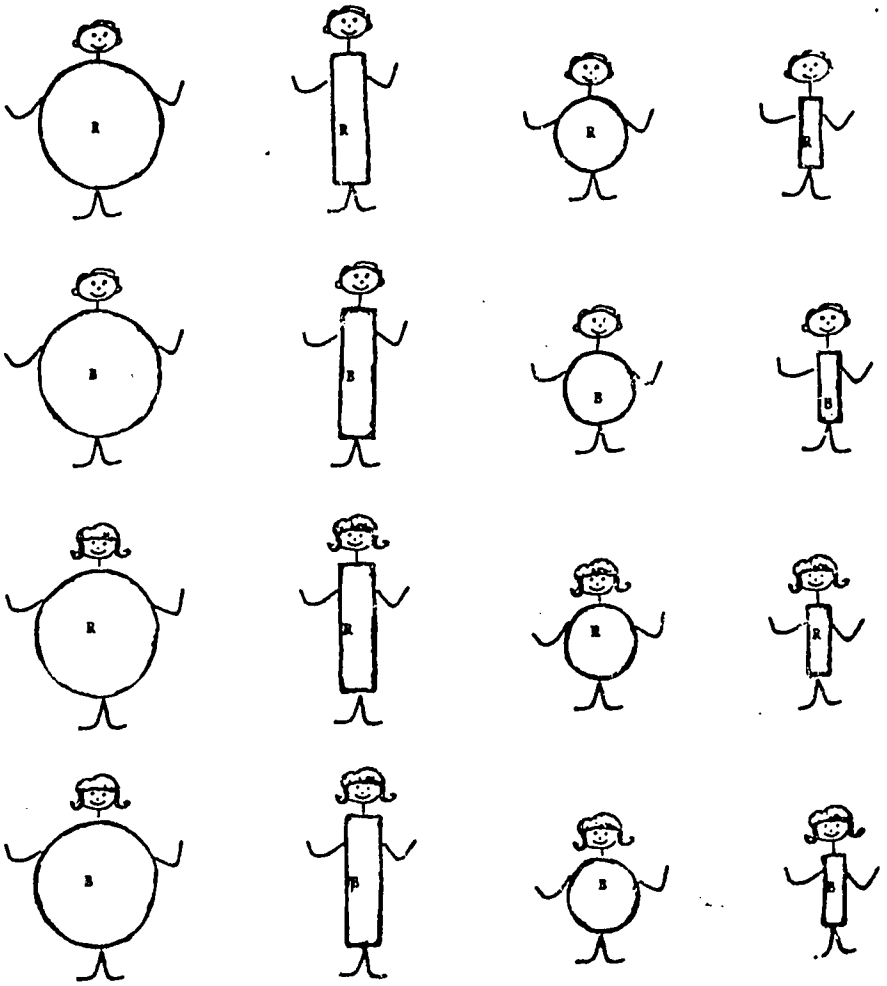


Figure 1. Picture of Children used in Test IIIa)
R: Colored Red, B: Colored Blue

Table 22

Item Analysis
Comparison Test IIIa): Classification
Items 1-10: Figures, Items 11-13: Shapes

Item		Percent Correct	
		CSMP	Non-CSMP
1. I'm thinking of one of these children but I'm not going to tell you which one. Find out as much as you can about the one I'm thinking about, but you can only ask three questions. They have to be questions I can answer with YES or NO. ("Target" was tall, thin, red girl.)	Attending to:		
	Height	43	38
	Width	74	86
	Color	67	60
	Sex	72	67
	Total	2.56	2.51*
	Non-Differentiable[1]	12	18
2. Now which one do you think it is? (the criterion was whether choice was consistent with responses in 1.)		31	28*
3. What can you tell me about this one (tall, thin, blue boy)? How is it different from the others?	Height	56	58
	Width	72	65
	Color	67	49
	Sex	62	68
	Total	2.57	2.40*
	Non-Differentiable[1]	23	45
4. What can you tell me about this one (point to short, fat, red girl)? How is it different from the others?	Height	52	50
	Width	89	82
	Color	54	49
	Sex	70	64
	Total	2.65	2.45*
	Non-Differentiable[1]	35	53
5. Find the short, thin, blue girl.		91	92*
6. Point to all the ones which are thin and red.	Two or more correct	99	99*
	All four correct	65	69*
7. Point to all the ones which are short and not fat.	Two or more correct	89	90*
	All four correct	80	79*
8. (After removing picture) How many red boys were there in the picture?	Response of 2,4 or 8	52	46*
	Response of 4	27	27*
9. (After putting picture back and covering the short, fat, blue girl) What can you tell me about the one that's missing?	Height	49	53
	Width	91	94
	Color	80	64
	Sex	94	90
	All four correct	16	17*
10. Look at this one (the short, fat, red girl). Which one is most different from it? In what ways is it different?	Height	43	37
	Width	85	81
	Color	48	41
	Sex	72	57
	Total	2.48	2.26*
11. (After showing student unsorted set of 26 shape cards.) Now one of these pieces is missing. Do you remember with the picture of the children you had to figure out which one was missing? Well this is the same thing.	Identified:		
	Color	68	79
	Shape	78	78
	Number of Dots	65	46
	Total	2.11	2.03*
*Mean Score (Sum of items with asterisk)		18.01	17.22
Standard Deviation		4.66	3.47
Correlation with Pretest		.53	.52
KR20		.74	.51

[1]Non-differentiable attributes refer to hair, eyes, arms and other characteristics which are in fact identical for the pictured children. Care had to be exercised when the characteristic, such as hair, was used to differentiate boys and girls, in which case it was a legitimate response. Responses dealing with rows or columns were given occasionally and were accepted.

As can be seen, CSMP students did slightly better on almost all the items though the differences were small in all cases. Whatever vocabulary the child used was accepted and continued after it was determined what was intended by the word (long, wide, round, square and many other words were used). On item 11, if the student did not get anything he was asked specifically what color the missing piece was, and then, if necessary, told the color and asked the shape. Thus each of the three items were in some sense independent. In assigning a total score partial credit was given in items 6, 7 and 8, and for item 9 credit was given only if the student gave all four characteristics of the missing piece.

Test IIIb) Binary Relations. For this subtest, a three dimensional representation like that below was shown to the students. A turtle sits on the railway track looking down the tracks in one direction or the other (i.e., towards either the mountain or the river). On his back sits a bug who also looks in one direction or another, though not necessarily the same as the turtle.

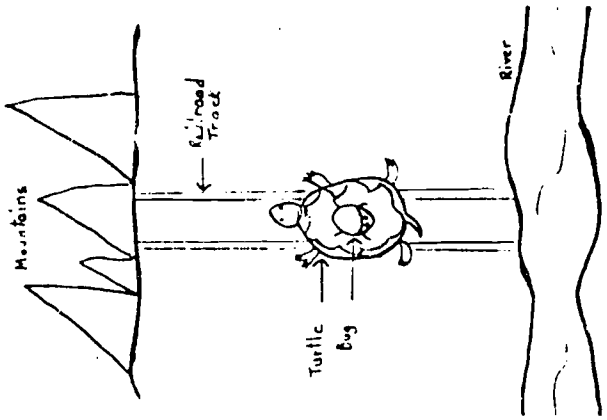


Figure 1. Demonstration Material for Test IIIb) (top view)

At the sound of a Train whistle (T for turtle) the turtle turns around and faces the opposite direction. The bug takes no action, but of course he too ends up facing the opposite direction. At the sound of a Bell (B for bug) the bug only turns around. The question then was to determine, for a given starting position, what the final positions of the bug and turtle would be after a particular series of train whistles and/or bells. With a little thought it can be seen that the turtle faces the same or opposite to his starting position according to whether the number of train whistles is even or odd respectively, and the bug faces the same or opposite according to whether the number of sounds (train whistles and bells) is even or odd and that the order of these sounds is immaterial. In the item analysis below, initial starting positions are given by "right" and "left", though on the actual test the realistic looking bug and turtle were actually placed to look at the mountain (left) or river (right). Students were not allowed to pick up the animals.

The materials were taken away for the last two questions, which were orally given problems involving a hypothetical flea crawling up a hypothetical flag pole. These items (see Table 23) touched on the idea of composition of functions.

Table 23

Item Analysis
Comparison Test IIIb): Binary Relations

Test Items				Percent Correct	
				CSMP	Non-CSMP
	Initial Position	Sounds	Final Position		
1.	Bug - right Turtle - left	T • B	Bug ? ? Turtle ?	89 99	95 97
2.	Bug - right Turtle - left	B • T	Bug ? Turtle ?	46 67	47 67
3.	Bug - right Turtle - left	T • T • T • B	Bug ? Turtle ?	47 62	53 72
4.	Bug - right Turtle - left	T • T • B • T • T • B	Bug ? Turtle ?	63 58	44 35
5.	Bug - right Turtle - left	10 train whistles then 5 bells	Bug ? Turtle ?	59 62	67 64
6.	Bug - right Turtle - left	What 3 sounds could produce { Bug - right, Turtle - left }	Bug ? Turtle ?	32	41
	(After taking away material) Suppose a flea is climbing up a tall flagpole. When he hears a bell he climbs up 3 inches, then on the next bell he slips back 2 inches, then up 3 inches, then back and so on.				
7.	How far up the flagpole will he be after 2 bells?			33	26
8.	How far up will he be after 100 bells?		Response < 100	77	79
			Response between 40 and 60	9	14
			Mean Score	8.01	8.00
			Standard Deviation	1.95	1.97
			Correlation with Pretest	.17	-.02
			KR20	.30	.34

The statistics for the two groups of students were almost identical. For some reason CSMP students did better on item 4, but Non-CSMP students made up for that over the rest of the items. The test was difficult, the responses to items 2 to 5 being near the guessing level, and the correlations and KR20 being very low.

Analysis of Class Means: The data in Table 24 are based on 12 pairs of classes, with from five to seven students in each class (selection of students is described on p. 7) taking the tests.

Table 24

Analysis of Covariance
Comparison Test III

	Mean Test Score Across Classes		Mean Pretest Score Across Classes		F-test, 1 and 10 degree of freedom p less than
	CSMP	Non-CSMP	CSMP	Non-CSMP	
IIIa) Classification	17.98	17.23	53.00	54.29	.71
IIIb) Binary Relations	8.05	8.12			.37
Total	26.03	25.36			.28

None of the differences were even close to significance. Using class means, the correlation between test and pretest range from -.23 to -.50 for the three scores, so that though CSMP classes had the lower pretest score, their mean test scores were actually adjusted downward while the mean scores for Non-CSMP classes were adjusted upwards. The adjusted mean total scores, for example, were 25.80 and 25.59 for CSMP and Non-CSMP classes respectively. The explanation for the differences between the positive correlations, when computed on an individual student basis, and the negative correlations, when computed on a class mean basis under the analysis of covariance model, is that within schools, i.e., class pairs, the class with the lower pretest score more often than not had a higher test score. Thus interpretation of this test is difficult. Perhaps the simplest summary statement is what has already been said - that there were no significant differences.

Comparison Test IV (Individually Administered)

The item statistics for these tests were based on 78 CSMP students with a mean pretest score of 53.14 and 74 Non-CSMP students with a mean pretest score of 54.35

Test IVa) Number Patterns: Students were shown a card, with a series of numbers and blank somewhere, such as 5, 6, 7, _, 9, and asked what could go in the blank. They were also asked to give the reason for their selection. If an unexpected answer was given they were asked to explain that answer and if they couldn't, to find out if any other number could go there. It was possible, though unusual, for a student to give a correct reason but to make an arithmetic error. Thus two parts were scored for each item.

Table 25

Item Analysis
Comparison Test IVa): Number Patterns

Test Item	Percent Correct*	
	CSMP	Non-CSMP
1. 5, 6, 7, __, 9	100 95	100 99
2. 2, 4, 6, 8, __	91 92	92 89
3. 19, 17, 15, __, 11	74 68	66 66
4. 15, 20, 25, 30, __	90 79	86 69
5. 1, 2, 4, 8, __	17 17	3 5
6. 16, 13, __, 7, 4	62 58	55 54
7. 0, 5, 1, 6, 2, __	29 22	28 23
8. 5, 9, 13, __, 21	63 55	55 50
Mean Score	10.11	9.42
Standard Deviation	3.52	3.53
Correlation with Pretest	.54	.62
KR20	.85	.86

*The second entry for each item is the percent who gave a correct reason for their choice of number.

CSMP students did better on almost all items, the differences usually being rather small. The largest differences occurred on item 4, which was essentially counting by fives. While these items probably test a rather general numerical ability, items similar to them often being found on tests of mental ability, CSMP students do considerable work with arrow diagrams where number patterns emerge in striking ways. They do not, however, do items such as these and therefore these results are encouraging.

Test IVb) Functions: For this test, the student was told that a teacher was playing a game with a class. Every time a student gave her a number the teacher did something, always the same thing, to the number and got a new number. The student was then shown a series of number-pairs on a card like the following:

4 changes to 7
11 changes to 14
5 changes to 8

This means that after the mysterious "something" was done to the numbers, 4 was changed to 7, 11 to 14 and 5 to 8. What was it that was done to each number? If a student gave no explanation or an incorrect one, and only then, he was asked "If we did the same thing to 7 what could we get?" Thus the student could get the item correct by explanation or, failing that, by example. If the student gave a correct but incomplete explanation, such as "made it larger" for the example above, he was encouraged to be more precise ("Yes that is correct, but they are larger in a special way").

For each series of number-pairs a second question was asked. This time the number the teacher ended up with was given and the question was what number must the student have given her. In the example above, the number that the teacher ended up with 4 (not shown) and therefore the number she started with must have been 1.

Table 26

Item Analysis
Comparison Test IVb): Functions

Number Pairs*	Percent Correct By Explanation		Percent Correct By Example		Percent Who Got Original Input	
	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP
1. 1-2, 6-7, 4-5, (16)	90	97	3	3	88	86
2. 4-7, 11-14, 5-8, (4)	86	82	3	7	67	57
3. 3-6, 5-10, 7-14, (12)	37	22	28	16	21	18
4. 27-25, 4-2, 15-13, (17)	78	74	0	3	33	31
5. 10-5, 2-1, 6-3, (6)	21	24	12	16	13	15
6. 2-6, 4-12, 1-3, (9)	13	3	3	1	10	1
7. 3-7, 2-5, 5-11, (13)	3	3	4	1	3	1
8. Johnny gave the teacher a number. She doubled it and then added one to it and ended up with 9. What was Johnny's number?					32	23
Mean Score					6.45	5.86
Standard Deviation					3.31	3.19
Correlation with Pretest					.50	.55
KR20					.81	.82

*The number in parentheses is the number given in the second part of the item, i.e., "What number did the student give the teacher for her to end up with this number?"

Again the differences are small but consistently in favor of CSMP students. This may have been a result of the considerable work done in the CSMP curriculum on return arrows.

Test IVc) Probability. In this test the student was to consider that painted blocks were rolled, like dice, to determine the color that would most likely appear on the top face after the roll. Several blocks were used, with different numbers of sides painted various colors. A block with two sides red, three sides blue and one side yellow is denoted by 2R-3B-1Y. Actual blocks with 1-inch sides were used and students were to pick them up and examine them but not to roll them.

Table 27

Item Analysis
Comparison Test IVc): Probability

Test Item	Percent Correct	
	CSMP	Non-CSMP
1. Given 5R-1B. What color is most likely to end up on top?	89	81
2. Why did you pick _____?	81	70
3. If you had some red and blue **paint**, how could you fix that block so that both red and blue have the same chance of ending up on top?	59	38
4. Given 2R-3B-1Y, 3R-2B-1Y, 4R-2B Which one is most likely to end up with blue on top?	72	64
5. Given 3R-2B-1Y. Suppose we tossed this block 12 times. About how many of those 12 times do you think red would end up on top?*	54	34
6. Suppose we tossed it 12 times again. About how many times do you think blue would end up on top?**	55	61
7. (Response for item 6 > response for item 5.)	50	51
Mean Score	4.60	4.08
Standard Deviation	1.50	1.62
Correlation with Pretest	.37	.19
KR20	.40	.47

*Response of 5,6 or 7 counted correct.
**Response of 1,2,3,4,5 counted as correct.

Again the differences were in favor of CSMP, particularly on items 3 and 5. CSMP students had some limited experience in probabilistic ideas, but they had not worked with blocks or dice, nor had they been given the notion of a probability as a number related to number of ways an event can occur.

Analysis of Class Means: The data in Table 28 are based on 11 pairs of CSMP - Non-CSMP classes. Usually about six or seven specially selected students from each class were administered these tests.

Table 28

Analysis of Covariance
Comparison Test IV

Test	Mean Test Score Across Classes		Mean Pretest Score Across Classes		F-test, 1 and 9 degrees of freedom p less than
	CSMP	Non-CSMP	CSMP	Non-CSMP	
Test IVa) Number Patterns	10.11	9.28	54.18	54.52	.05
Test IVb) Functions	6.48	5.75			.10
Test IVc) Probability	4.59	3.99			.05
Total Test IV	21.18	19.01			.02

It can be seen that all differences were in favor of CSMP and all were significant except Test IVb): Functions, which nevertheless approached significance. The adjustment of mean scores due to pretest differences was small (<.04) in all cases. The phenomenon of positive correlations between pretest and test scores when based on students and negative correlations when based on classes recurred for Test IVc), the correlation by classes being -.54.

CSMP Tests

Construction of Tests

The second grade CSMP curriculum consists largely of a detailed set of lesson plans, complemented by a series of workbooks and worksheets. There are 287 lessons which use a spiral development where a child is introduced to each of several topics, a little at a time and then proceeds through increasing levels of sophistication with each topic. In each of the eight workbook series which complement the lessons, there are seven workbooks graduated in difficulty from the first which is almost remedial in nature to the seventh which is difficult and intended only for the best student.

Generally only those topics which had been covered extensively and with which the children had had considerable practice were considered for inclusion in the end-of-year test.

The basic source of items was the workbook series as it was felt the problems in the workbooks provide the most accurate indication of what was expected of the students. Since many of the teachers did not complete all eight of the workbook series, only the topics covered in the first five workbook series (about 220 lessons) were considered. Furthermore all of the content selected from the workbooks was from the first five levels of each workbook series.

The first 220 lessons were analyzed to ensure the content selected from the workbooks reflected and included the content of the lessons. A list of topics included in the lessons is given below and the number of lessons related to that topic is given in parenthesis. About half of the lessons were concerned with more than one topic, hence the total number of lessons shown greatly exceeds 287.

Table 29

Topics In CSMP
Second Grade Curriculum

Topic	Number of Lessons
Addition	93
Subtraction	61
Multiplication	53
Integers	30
Rationals	31
Minicomputer	50
Relations and Arrow Diagrams	59
Open Sentences	18
Classification-Sets and String Diagrams	20
Parenthesis	18
Combinatorics	19
Probability	7
Numerals and Counting	12
Geometry and Measurement	17
Mental Arithmetic	18
Order	10

Some of the above topics were covered at least partially on the Standardized Test; others, while not tested directly on the Comparison Tests, were at least related (in some cases rather vaguely) to one or more of those subtests. These topics were not repeated; hence the CSMP tests which were finally constructed tended to contain only material not found in other curricula (such as the use of the Minicomputer and Arrow Diagrams) or not usually found at the second grade level (such as integers or multiplication by larger numbers). One certainly cannot look on the tests in this chapter as an operational definition of the CSMP curriculum; they show, in a sense, the residuals of the tests in the previous two chapters.

Four test booklets (A to D) were constructed and administered to eight or nine classes according to the testing plan given on page 7. Each test usually required about 20 minutes but was essentially untimed as students worked through the test at their own speed. The format of both the test booklets and of the test items closely resembled that of the workbooks. Students received help ("What to do" not "how to do it") from the test administrators whenever necessary, usually on an individual basis. The first draft of each test was pilot tested in three second grade CSMP classes which were not scheduled in the testing plan to take the particular test.

A brief listing of the content tested on the four tests is given below.

Test A: Open sentences and parentheses (6 items)
Whole number multiplication (10 items)
Arrow diagrams (5 items)
Test B: Fractional parts of whole numbers (10 items)
Interpreting an arrow diagram (4 items)
Composition of functions: arrow diagrams (3 items)
Test C: Integers: ordering and adding (20 items)
Multiplication - ordering (7 items)
Test D: "Solution" of arrow diagrams: see item analysis (5 problems)
In addition Tests A, B and C each had a four-item problem on the converse of a relation (return arrows).

A fifth test, denoted Test E, was administered to a representative sample of six students from each class of the classes shown in the testing plan. This was an individually administered test of students' ability to use the Minicomputer.

Results

The results will be presented in the form of item analysis. Within each test items were grouped together logically to make up subtests. The order in which the subtests are given here was chosen for ease of interpretation and does not correspond to the order of the tests themselves. The statistics were derived by averaging across all students who took the test and this includes data from classes 14, 15 and 16, which classes have not been considered in previous analyses since there were no appropriate comparison classes available for them. In the previous two chapters, all students who were in the class in the fall and present for the pretest were considered for both item analyses and class means. In this chapter only students who were in CSMP first grade last year (for at least the last four months) and who were present for the pretest are considered. Differences between these students and those who started CSMP in second grade will be considered for all tests in the next chapter.

The mean pretest scores for students who took tests A, B, C and D respectively were 50.7, 52.3, 50.6 and 50.3, and correlations between the pretest and each of the tests were also remarkably similar: .70, .64, .68, and .60 respectively.

Table 30

Item Analysis
CSMP Subtest 1

Test Items (from Test A)	Percent Correct
1. 2 x 6 = ___	83
2. 8 x 0 = ___	64
3. 3 x 3 = ___	86
4. 2 x 13 = ___	67
5. 3 x 12 = ___	49
6. 2 x 37 = ___	44
Mean Score	3.93
Correlation with Pretest	.64
KR20	.78

The percent correct for items 1-3, Table 30, is in the same range as for similar items on the Standardized Test. The percents correct for items 4-6, not covered on standardized tests at this grade level, are lower, but still quite high for items of this nature. Particularly impressive, to the author at least, is the fact that 44% of the students could calculate "2x37" correctly.

Table 31

Item Analysis
CSMP Subtest 2

Test Items (from Test A)	Percent Correct
1. 2 X 37 = ___ (30+7)	40
2. 2 X 45 = ___ (+)	38
3. 2 X 13 = ___	60
4. 3 X 14 = ___	38
Mean Score	1.76
Correlation with Pretest	.55
KR20	.86

The items in Table 31 are similar to those in Table 30. The difference is that the items in Table 30 appeared first in the test; later came the items in Table 31. These latter items included a clue for the students; a reminder of a way in which they had been taught to do these problems. Items 4 and 6 of Table 30 are in fact the same as items 3 and 1 respectively of Table 31. Contrary to expectation, students did not do better when given a clue; in fact percents correct were slightly lower. Not unexpectedly, 38% of the students got none of these four items correct.

Table 32

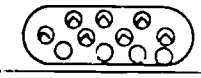
Item Analysis
CSMP Subtest 3

Test Items (from Test C); Put <, = or > between each pair			Percent Correct
1.	4x6	6x4	69
2.	5x22	5x25	90
3.	4x(3+2)	4x5	49
4.	$\frac{1}{2}$x12	$\frac{1}{3}$x12	63
5.	$\frac{1}{2}$x58	$\frac{1}{2}$x56	79
6.	$\frac{1}{4}$x27	$\frac{1}{2}$x27	58
7.	$\frac{1}{5}$x240	$\frac{2}{5}$x240	40
	Mean Score		4.48
	Correlation with Pretest		.61
	KR20		.40

The items in Table 32 tested what might be called basic understanding of the process of multiplication; no computation was required. The last four items should be reviewed in the context of Table 6, which deals with calculating halves and thirds of whole numbers. Student performance seems adequate, though one might have thought that more students would have known that $\frac{1}{2}$ of 12 is greater than $\frac{1}{3}$ of 12. The relatively low value of KR20, indicating that the test items were not particularly homogeneous, may be due to the two kinds of items involving whole numbers and fractions, and to the element of guessing.

Table 33

Item Analysis
CSMP Subtest 4

Test Items (from Test C)	Percent Correct
1. $60 + \hat{20} = \hat{80}$ $60 + \hat{20} = \hat{40}$ $60 + \hat{20} = 40$ $60 + \hat{20} = 80$ Circle the true number sentence	62
2. $25 + \hat{35} = 60$ $25 + \hat{35} = \hat{10}$ $25 + \hat{35} = 10$ $25 + \hat{35} = 60$ Circle the true number sentence	42
3. $8 + \hat{6} =$	71
4. $4 + \hat{7} =$	52
5. $\hat{5} + \hat{4} =$	52
6. $\hat{14} + 3 =$	42
7. $\hat{5} + 5 =$	68
8. $\hat{15} + \hat{4} =$	57
9. $\hat{5} + 10 =$	61
10. $4 + \hat{7} =$ 	81
11. $\hat{5} + 11 =$ 	74
12. $\hat{7} + \hat{4} =$	49
13. $\hat{5} + 10 =$	79
14. $4 + \hat{16} =$	53
Mean Score	8.43
Correlation with Pretest	.63
KR20	.87

The items in Table 33 deal with adding integers, including negative numbers.

A perusal of Table 33 indicates that all four of the items with a positive answer (items 3, 9, 11 and 14) were more frequently correct than all seven of the items with a negative answer (items 4-6, 8, 10, 12 and 14). When clues were given, exactly as shown, for the page containing items 10-14 performance improved. With the addition "$4+\hat{7}=$" percent correct rose from 52% (item 4) to 81% (item 10) and on "$\hat{5}+10$" from 61% (item 9) to 79% (item 13). Student responses indicate that about half of the incorrect answers could be accounted for by one of three explanations: a) the answer was reversed ($8+\hat{6}$ would be $\hat{2}$), b) a "hat" was ignored

($8+\hat{6}$ would be 14), or c) a "hat" was added ($8+\hat{6}$ would be $\hat{14}$). These three errors occurred about equally often.

Integers are not taught in other second grade curricula so there is no basis for comparison of performance. There was only a slight improvement from first grade. For example 61% of last year's first graders got $3+\hat{8}$ and 67% of this year's second graders (generally the same students) got $5+\hat{5}$. Last year 44% got $2+\hat{3}$ and this year 52% got $4+\hat{7}$. Last year 61% got $9+\hat{1}$ and this year 71% got $8+\hat{6}$.

Table 34

Item Analysis
CSMP Subtest 5

Test Items (from Test C) Put <, = or > between each pair			Percent Correct Percent
1.	5	$\hat{5}$	76
2.	$\hat{12}$	0	51
3.	$\hat{6}$	$\hat{8}$	35
4.	$\hat{12+8}$	$\hat{14+8}$	32
5.	$\hat{58+49}$	$\hat{58+45}$	66
6.	$63+\hat{45}$	$\hat{63+45}$	51
	Mean Score		3.11
	Correlation with Pretest		.37
	KR20		.52

Students had the most difficulty with items 3 and 4. These were the only two items in which the two numbers to be compared were both negative. Probably many students compared the absolute values and obtained, for example, $\hat{8}<\hat{6}$ (item 3).

Table 35

Item Analysis
CSMP Subtest 6

Test Items (from Test B)	Percent Correct
1. $\frac{1}{2}$ x 6 =	94
2. $\frac{1}{3}$ x 18 =	60
3. $\frac{1}{3}$ x 9 =	81
4. $\frac{1}{2}$ x 48 =	63
5. $\frac{1}{2}$ x 14 = (△ + △ = 14)	81
6. $\frac{1}{3}$ x 18 =	62
7. $\frac{1}{4}$ x 20 =	57
8. $\frac{1}{2}$ x 48 = (40 + 8)	71
9. $\frac{1}{3}$ x 63 = (__ + __)	33
10. $\frac{1}{2}$ x 254 =	7
Mean Score	6.08
Correlation with Pretest	.57
KR20	.86

This topic, multiplying by halves and thirds, also does not appear on standardized tests at this grade level. The performance of CSMP students is impressive. Items 5-7, with one kind of clue appeared on a later test page than did items 1-4. Thus one can compare performance on $\frac{1}{3}$x18 from item 2 (without clue) and item 6 (with clue). Evidently the clue did not help. Similarly one can compare $\frac{1}{2}$x48 from item 4 and item 8 and see only a moderate improvement.

Table 36

Item Analysis
CSMP Subtest 7

Test Items (from Test A)	Percent Correct
Put numbers in the boxes to make the sentences true	
1. $9 - \square = 6$	91
2. $\frac{1}{2} \times \square = 4$	67
3. $\square + 5 = 7$	93
4. $3 \times \square = 12$	64
Mean Score	3.16
Correlation with Pretest	.50
KR20	.57

The items in Table 36 tested students' ability to work with open sentences and the results were quite good. Although items 2 and 4 were more difficult, as expected, one is impressed by the fact that two-thirds of the students did get them correct. The moderate value of KR20 indicates that the items were not particularly homogeneous, probably because the various computational skills required were more important for success than any generalized understanding of how to work with open sentences.

Table 37

Item Analysis
CSMP Subtest 8

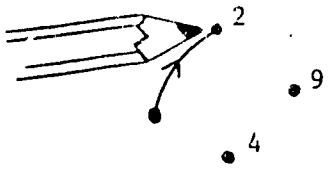
Test Items (from Test A)	Percent Correct
1. $6 + (4 - 2) =$	60
2. $(2 + 2) + (3 - 1) =$	59
Mean Score	1.19
Correlation with Pretest	.44
KR20	.64

There was an error in printing item 2 from Table 37. The middle sign should have been minus, yielding an answer of 4-2=2. Then item 1 would have been correct whether or not the students ignored the parenthesis but for item 2 the student would have had to work first within the parenthesis. Otherwise his answer of 2+2-3-1 would have been wrong. What might have been expected then would be a high percent correct on item 1 and some lower percent correct for item 2. The percent correct on the actual items was only moderate, however. One suspects, based on student interviews* that many students put down as their answer the contents on the parenthesis, rather than the complete expression.

*Evaluation Report 2-B-3: Student Interviews

Table 38

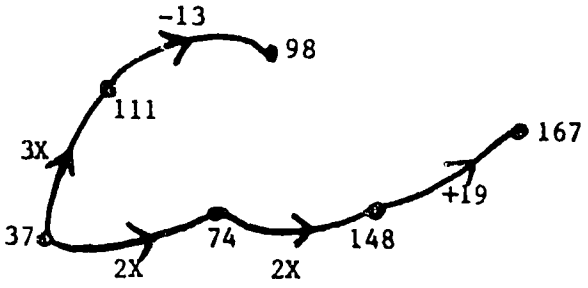
Item Analysis
CSMP Subtest 9

Test Items (From Test A)	Percent Correct
Draw arrows for "I am more than you" 	
1. Arrow from 9 to 7	74
2. Arrow from 9 to 4	84
3. Arrow from 9 to 2	81
4. Arrow from 7 to 4	81
5. Arrow from 4 to 2	80
Mean Score	4.01
Correlation with Pretest	.43
KR20	.90

The item in Table 38 was repeated from a test administered in first grade to these same students. The mean number correct increased from 3.1 to 4.0 during that time. The mean number correct for the lowest scoring quarter increased from 0.95 to 1.35; nevertheless 15% of the students failed to get more than one of the 5 arrows drawn correctly. Sixty nine percent of the students got all arrows drawn correctly.

Table 39

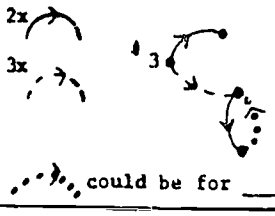
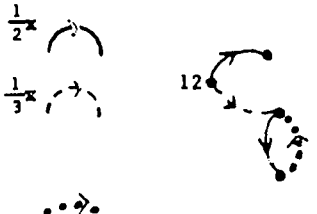
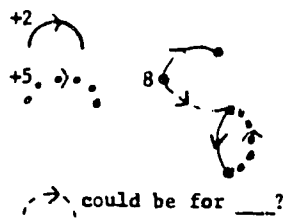
Item Analysis
CSMP Subtest 10

Test Items (from Test B)*	Percent Correct
	
1. 3 x 37 =	64
2. $\frac{1}{2}$ x 74 =	37
3. 148 + 19 =	57
4. 4 x 37 =	29
Mean Score	1.97
Correlation with Pretest	.51
KR20	.71

*A different color was used for each different arrow.

The diagram in Table 39 is an example of what is sometimes called a "Detective Story" in the CSMP curriculum. There is a mass of information given which was slightly less confusing with actual colors. Some answers are given explicitly as in tracking down the answers to items 1 and 3 by searching out the required information which is given directly for the student must realize that is 2x37=74 (shown) then $\frac{1}{2}$x74=37. This is the idea of the "return arrow" ($\frac{1}{2}$x) from 74 to 37. Similarly in item 4, a 2x arrow followed by a 2x arrow can be replaced by what we might call the composite 4x arrow. It is difficult to evaluate the adequacy of performance on this subtest. Although students do get considerable practice with return arrows and ideas of composition, there are rather infrequent lessons dealing with diagrams as "packed" as this one and they are clearly intended as difficult problems. About a quarter of the students did not get any items correct and it may be that for many of them the amount of information was simply overwhelming.

Table 40
Item Analysis
CSMP Subtests 11-13

Test Item (Label dots and arrows)	Percent Correct
1. - 4. Subtest 11 (from Test A)* 	64 62 52 33
5. - 8. Subtest 12 (from Test B)* 	72 56 58 51
9. - 12. Subtest 13 (from Test C)* 	51 54 53 36
Range of Mean Scores	1.95 - 2.04
Range of Correlations with Pretest	.50 - .61
Range of KR 20's	.79 - .83

*Different colors were used for each different arrow.

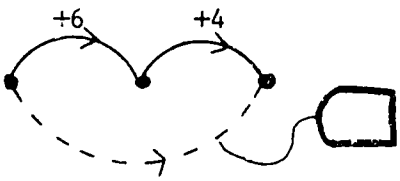
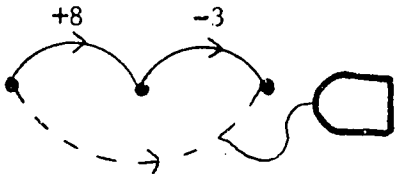
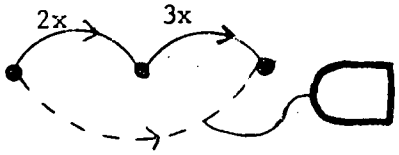
Subtests 11-13 were identically formatted problems involving different operations. Three parts of each problem dealt with labelling dots on a fairly straightforward diagram. The computations were not difficult; for items 1-3, 5, 6 and 10 there were similar or identical computations elsewhere in the CSMP or Standardized tests. (For example the computation 3x3 is required for item 2 and that computation also appears in Subtest 1.) For these six items, the mean percent correct when the item was posed in standard form without arrows was 79, and for the same items in the context of an arrow diagram the percent correct was only 62. Thus there appears to be a significant number of students who were unable to figure out what to do - even though the diagrams are not particularly complicated (they were much simpler looking in color). For each of the three problems nearly a third of the students did not get any answers correct.

There were two ways of determining what the answers to the last question of each section, what the answers could be for. One was to look at the arrow as a return arrow; knowing what the dots are for is unnecessary. The other was to look at the dots to determine what arrow could work between them, thus disregarding the opposite arrow. In the first problem, correct answers were about evenly divided between 2x (probably the first method) and +2 (probably the second method) and 29% of the students gave the incorrect answer of 4x, evidently a composition of two 2x arrows. In the second problem, correct answers were again fairly evenly divided between $\frac{1}{2}x$ and -9, and 28% of the students gave for their incorrect answer the

composition of two $\frac{1}{2}x$ arrows, i.e., $\frac{1}{4}x$. In the third problem almost all correct answers were +2, expected under either method of solution, but there were no particularly frequent incorrect errors. In fact there was a bewildering array of them.

Table 41

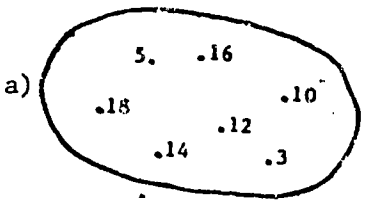
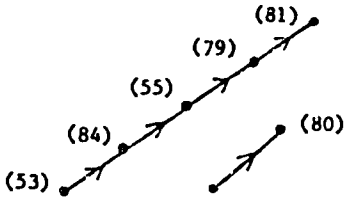
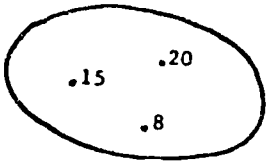
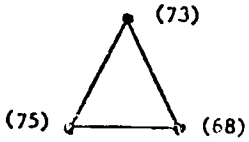
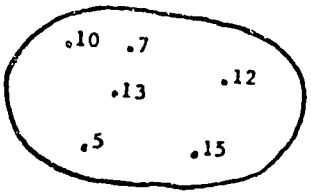
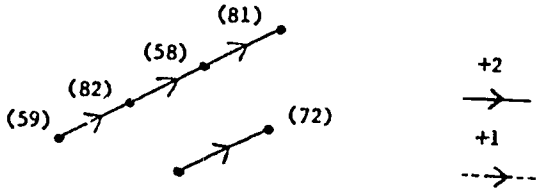
Item Analysis
CSMP Subtest 14

Test Item (from Test B): Label the arrows	Percent Correct (Sign omitted)
1. 	43 (19)
2. 	37 (16)
3. 	3 (6)
Mean Score	0.83 (.41)
Correlation with Pretest	.46
KR20	.60

The items in Table 41 dealt exclusively with composite arrows. Students were told that they could label dots to help them if they wanted to. Forty percent of the students did this, usually for all three items. The figures in parenthesis are the percent of students, in addition to those who got the correct answer, who left the operation off their answer. (In the first item this would give an answer of "10" instead of "+10".) They were not counted as correct since 10 is also the label for the last dot if one starts from what might be a natural starting point of zero. This method works for the first two items (from 0 to 10 is +10 and from 0 to 5 is +5) but would yield an answer of (+?) 0 (from 0 to 0) or just plain confusion for item 3. These were difficult items; over half the students got none of the three correct.

Table 42

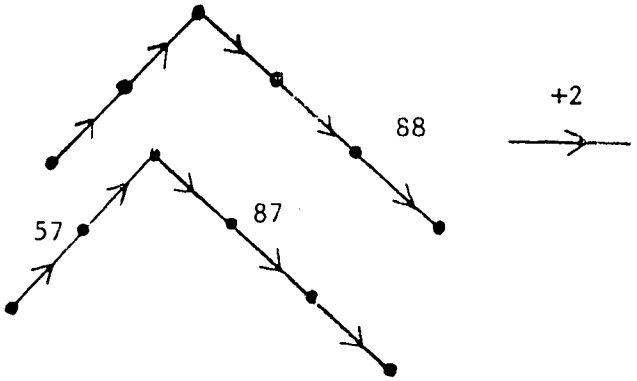
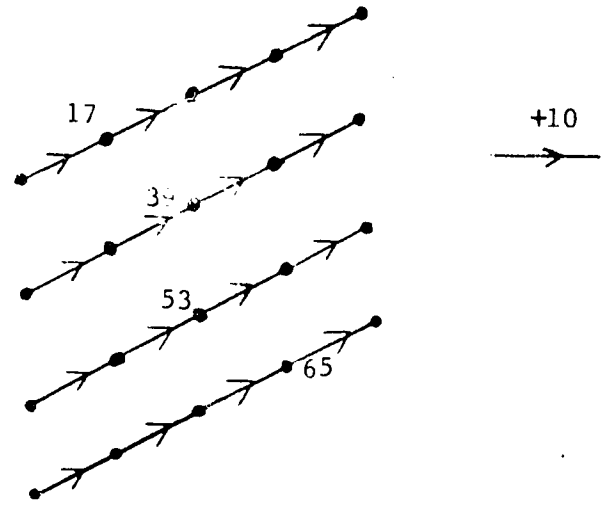
Item Analysis
CSMP Subtest 15 (from Test D)

Match these numbers	to these dots (Dots were not numbered; see below for interpretation)	Percent All Correct
a) 		48
b) 		64
c) 		51
Range of Correlations with Pretest		.21 - .43
Range of KR20's		.83 - .88

These problems were based on a format being tried out by the developers of the curriculum. Thus these students had not seen this type of problem before and considerable explanation, with examples and individual assistance, was required. The idea in each problem was to use each of the numbers given by the string on the left exactly once on the dots in the diagram to the right. The dots were not labelled - only the code for the arrow was given. Thus, the problem had a unique solution and there was much backtracking and figuring out exactly how one should go about starting the problem. The numbers in parenthesis are the number who got the right number for the dot, in the sense that the number was consistent with the numbers given on adjacent dot(s) according to the arrow joining them. The exception in items 1 and 3 was the left most dot of the longer path. That had to be the correct answer for the problem to be solved. Thus it can be seen that almost all students who got the correct number for that dot also got all the other dots labelled. The most common error was to start at the left most dot with the wrong "low number". In item a), either the long path would be 3, 5, 10 (wrong and the percent correct was lower), 12, 14 and the short one 16, 18; or else 3, 5, 14 (wrong again), 16, 18 and 10, 12. Thus the percent correct for the first and third dot in the long path is distinctly lower than for the remaining dots, which are all about the same. The same pattern is apparent in item c), where students would start the long path with 3, 5 then get stuck and finish with 13, 15 or 10, 12.

Table 43

Item Analysis
CSMP Subtest 16

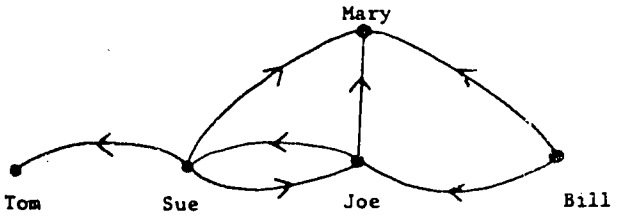
Which dot is for 83 (from Test D)	Percent Correct
a) 	57
b) 	52

Mean Score	1.10
Correlation with Pretest	.54
KR20	.69

The format of these two problems was also unfamiliar to students. Each problem was solved by just over half the students. Students were told they could label as many dots as they wanted to, but that all they needed to do was to find and label the dot which was for 83. It is not possible to summarize the wide variety of responses for these items. More than half of the students who got the items correct labelled only one dot, 83, evidently figuring out the problem mentally. About 15 percent of the students labelled five or more dots incorrectly.

Test 44

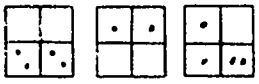
Item Analysis
CSMP Subtest 17

Test Items (from Test D)	Percent Correct
The arrows are for "I sent a card to you" 	
1. Who got the most cards? _____	78
2. Who did not get any cards? _____	73
3. Who got exactly 2 cards? _____	67
4. Who sent the most cards? _____	55
Mean Score	2.73
Correlation with Pretest	.52
KR20	.77

The problem in Table 44 was to interpret a non-numeric arrow diagram. The reader may verify that it is surprisingly easy to make an error on any of these items. The decreasing percent correct may be partially explained by a regular increase in the percent of students who did not put down any answer at all (5%, 8%, 12%, then 14%), this being the last subtest of Test D. On only one item did more than nine percent of the students answer any particular wrong answer. On item 4, 14 percent of the students said Bill sent the most cards; in fact he only tied with Joe for second most. Twelve percent of the students got none of the four items right and 45 percent got all four correct.

Minicomputer Test (Test E)

This test was administered individually to students in nine classes (usually six per class). The tester took the student through the series of Minicomputer tasks given below.

1. What number is shown on the Minicomputer? 

   (a) Does the student know enough to make a forward play (i.e., to reduce to a standard configuration)?
   (b) Does the student make the backwards 40=20+20 move, in order to play 80+20=100?
   (c) Does the student make the backwards 4=2+2 move in order to play 8+2=10?
   (d) (If not - remove two ones checkers). Does the student make 8+2=10 move?
   (e) Does the student read answer correctly?

2. 427+218

   (a) Does the student set up Minicomputer correctly?
   (b) Does the student begin making forward plays?

3. 2x305

   (a) Does the student set up Minicomputer correctly?
   (b) Does the student begin making forward plays?
   (c) Can he read answer correctly?

4. 64-9

   (a) Does the student set up Minicomputer correctly?
   (b) Does the student begin to make backward moves?
   (c) Does the student cancel at least one pair of checkers?
   (d) Does the student make necessary 10=8+2 backwards move?
   (e) Does the student get correct answer?
   (f) Does the student read answer correctly?

5. 768-427

   (a) Does the student set up Minicomputer correctly?
   (b) Does the student begin to make backward moves?
   (c) Does the student cancel at least one pair of checkers?
   (d) Does the student get correct answer?
   (e) Does the student read answer correctly?

6. $\frac{1}{2}$x320

   (a) Does the student set up Minicomputer correctly?
   (b) Does the student make a backward move to get pair of checkers?
   (c) Does the student make particular backwards move 100=80+20?
   (d) Does the student get correct answer?
   (e) Does the student read answer correctly?

7. 3x24

   (a) Does the student set up Minicomputer correctly?
   (b) Does the student start to make plays?

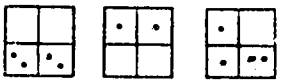
8. $\frac{1}{3}$x15

   (a) Does the student set up Minicomputer correctly?
   (b) Does the student make a backward move to get pair of checkers?
   (c) Does the student make particular backwards move 10=8+2?
   (d) Does the student get correct answer?
   (e) Does the student read answer correctly?

9. $\frac{1}{2}$x21

   (a) Does the student set up Minicomputer correctly?
   (b) Does the student make backward move 20=10+10?
   (c) Does the student get correct answer?

10. Given the following configuration:



    Does the number get larger, smaller or still the same if this is done?

    (a) Move checker on 80 to 40.
    (b) Take two checkers off 100 and put one checker on 200.
    (c) Move checker on 2 to 20.

    (a) - c) were demonstrated to student without comment.)

Considerable leeway was taken in administering and coding the test. The emphasis was not on whether the student could get the correct answer but on whether he knew how to go about doing the problem and whether he had certain rather general "Minicomputer Skills." Students were not necessarily given every item. For example when the student was clearly unable to do basic things he was not given the later more difficult problems or if the student had already, on item 5, demonstrated skill in subtracting on the Minicomputer then item 6 was omitted. Occasionally the tester was uncertain about the student's understanding and certain kinds of questions might be repeated.

A set of 12 Minicomputer skills was considered and a determination made as to whether or not the student possessed this skill (or knowledge as the case may be). That is, a binary (Yes - No) decision was made. The student was given the benefit of the doubt; for example if he once demonstrated how to set up the Minicomputer and correctly began to subtract, subsequent errors would not change the positive decision that he did know how to subtract on the Minicomputer. In most cases this decision was relatively easy to make - it was apparent to the tester. For some of the skills (9, 10, 12 below) it is possible, though unlikely, that the student might do much better or worse on another day. Skill 5 was difficult to judge, primarily because of guessing, and probably has low reliability. In parenthesis after each skill is given the percent of students who were judged to possess that

skill. The skills are given in order of observed difficulty.

1. Read numbers shown in standard form on the Minicomputer (81%)
2. Show given numbers on the Minicomputer (81%)
3. Make simple forward plays on the Minicomputer (79%)
4. Knows how to add numbers on the Minicomputer (77%)
5. Conservation on the Minicomputer (at least two correct from #10 above) (70%)
6. Knows to make "special" forward play (8+2=10) when simplifying configuration in adding or multiplying (68%)
7. Knows how to multiply numbers on the Minicomputer (49%)
8. Knows how to subtract on the Minicomputer (49%)
9. Knows to make "special" backwards play to ones board (10=8+2) when required for subtraction or division (45%)
10. Knows to play 4=2+2 in order to produce necessary 8+2 configuration when simplifying on Minicomputer (40%)
11. Knows how to use Minicomputer to find $\frac{1}{2}$, $\frac{1}{3}$ of a number (30%)
12. Knows to make special backward plays to tens board (100=80+20) when required for subtraction or division (23%)

Basically the list collapses to three categories: reading and writing on the Minicomputer (1 and 2), how to set up and begin to solve an arithmetic operation (4,7,8,11), and techniques necessary to complete the solution (3,6,9,11). The test (using these 12 skills to generate a test score) was very homogeneous (KR=.91) and highly correlated with the pretest (r=.68). With the exception of skill 5, the skills were highly ordered. For instance 15 students did not have all four of the first four skills, and only five of these 15 had even one (none more than one) of skills 6-12. Of the 28 students who did not get at least seven of the first eight skills (again excluding skill 5), only three got skill 10 and none got skills 11 and 12.

There was a wide diversity of scores. Low ability students (as measured by pretest scores) did very poorly. Of the eleven students with the lowest pretest scores (about 20% of the students tested), only two could both read and show numbers on the Minicomputer and only one of the eleven had more than four of the 12 skills. On the other hand 10 students were correct on everything and another seven missed at most two skills. The 53 students who were tested on the Minicomputer were selected to representative of local second grade CSMP students and all have been in the program since the beginning of the first grade.

Further Analysis of Test Data

Consistency of Results Across School Districts

In order to determine the degree of uniformity of these results across the four participating school districts, mean scores across classes were calculated for each district. This information is given in Table 45. Since different classes often wrote different tests (see the Testing Plan, p. 7) the total score given in the table is not based on any single group of students. It can be seen that fairly similar results were obtained in each district. However, these scores do not take into account differences in entering ability. Table 46 presents the same data, except that prior to calculating mean scores by district the mean score for each class was adjusted by the usual procedure* to take these differences into account. With this adjustment the results are seen to be quite consistent, particularly in the comparison of CSMP - Non-CSMP performance, across the various districts. Thus it seems unlikely that results found previously significant can be attributed to particularly favorable conditions in only one or two of the four districts.

Table 45

Mean Scores By District

Tests	District 1		District 2		District 3		District 4	
	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP
Comparison Tests								
I	18.4	16.6	20.2	18.9	21.2	16.9	14.1	12.0
II	37.8	31.2	46.1	42.6	41.3	36.4	32.3	34.4
III	26.4	26.3	27.5	25.7	25.7	25.3	22.5	23.6
IV	22.2	20.2	22.2	20.7	20.7	18.2	18.9	17.0
Standardized Tests								
Computation	20.1	19.4	23.2	21.7	22.2	20.4	19.5	18.4
Concepts	18.6	18.3	20.0	19.8	20.2	17.6	16.2	15.8
Total of Above	143.5	132.0	159.2	149.4	151.3	134.8	123.5	121.2

Table 46

Adjusted* Mean Scores by District

Tests	District 1		District 2		District 3		District 4	
	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP
Comparison Tests								
I	16.9	15.3	18.6	16.3	19.2	16.8	21.7	21.9
II	39.2	33.5	41.2	38.5	40.0	36.8	37.5	37.3
III	26.8	25.5	27.8	27.2	25.0	25.6	21.2	22.8
IV	22.2	20.7	21.7	19.4	20.8	18.5	19.7	17.5
Standardized Tests								
Computation	19.8	19.1	21.4	19.0	20.7	20.3	24.2	23.4
Concepts	18.0	17.9	17.0	15.6	17.8	17.6	23.8	23.2
Total of Above	142.9	132.0	147.7	136.0	143.5	135.6	148.1	144.1

*Adjusted class mean = class mean - r(class pretest score - mean pretest score), where r is the correlation between test and pretest scores.

Students Who transfer Into CSMP

An important question about the CSMP curriculum is the degree to which students can successfully transfer into and out of the CSMP program from a traditional math program. It has not been possible to track individual students who leave the program; however, the very strong performance of CSMP students on the standardized tests is solid evidence that these students would be well prepared, cognitively, for entering a traditional math program.

The problem for students entering the program, however, is much different. The languages of the Minicomputer, of arrow diagrams and of string pictures do not occur in traditional programs and so clearly these students have some catching up to do. Some students transferred into a second grade CSMP class at or near the beginning of the year and it has been possible to investigate the performance of such students. Their test scores were analyzed separately, as were the scores for students who were in the program from before the middle of first grade. In addition to computing mean test and pretest scores, a regression equation was developed for each group so that mean test scores for students of various ability groups could be estimated. Comparison Tests III and IV had a very small number of "late enterers" and this analysis was not done for them.

Table 47

Test Scores By Time of Entry:
(1)Before Feb. of First Grade (2)During Sept. or Oct. of Second Grade

		Number of Students	Mean Test Scores	Mean Pretest Scores	Estimated Test Score for Indicated Pretest Scores* 40	50	60
Comparison Test 1	Entry(1)	85	19.4	50.1	15.5	19.4	23.2
	Entry(2)	24	17.1	46.8			
Comparison Test 2	Entry(1)	111	38.1	53.0	30.4	36.3	42.3
	Entry(2)	26	36.6	48.9	31.1	37.3	43.4
CTBS: Computation	Entry(1)	198	22.2	51.8	18.2	21.0	23.7
	Entry(2)	50	20.8	48.1	17.7	20.7	23.7
CTBS: Concepts and Applications	Entry(1)	203	19.4	51.9	16.4	18.9	21.5
	Entry(2)	51	17.8	48.2	15.5	18.3	21.1
Total of Above Tests	Entry(1)				80.5	95.6	110.7
	Entry(2)				79.0	94.5	110.0
CSMP Test A	Entry(1)	63	17.9	53.2	13.4	16.8	20.2
	Entry(2)	17	14.2	44.2	12.9	16.1	19.2
CSMP Test B	Entry(1)	83	10.4	51.5	7.1	10.0	12.8
	Entry(2)	21	9.2	53.0	4.9	8.2	19.2
CSMP Test C	Entry(1)	102	18.6	51.9	14.3	17.9	21.5
	Entry(2)	27	13.4	53.0	11.2	14.5	17.9
CSMP Test D	Entry(1)	103	14.5	51.1	11.5	14.2	16.8
	Entry(2)	31	10.5	49.8	7.4	10.5	13.6
Total of CSMP Tests	Entry(1)				46.3	58.9	71.3
	Entry(2)				36.4	49.3	62.3

*Obtained substituting 40, 50 and 60 respectively for x in the equation $y' = \bar{y} + r\frac{s_y}{s_x}(x - \bar{x})$ where $\bar{x}$ and $\bar{y}$ are the group mean scores for pretest and test respectively, s_x, s_y are the corresponding standard deviations and r is the correlation coefficient.

The first four tests listed in Table 47 are tests written by both CSMP and Non-CSMP students and do not require the concepts and skills unique to CSMP. The group of students who transferred into the classes usually had lower pretest scores and lower test scores. However, when regression estimates were made for students with mean pretest scores of 40, 50 and 60 (roughly the 20th, 50th and 75th percentile among all the students who were pretested), it is found that whether the student joined their class in first grade or in second grade made very little difference in his scores. The data suggests that CSMP students are not penalized in the acquisition of standard content by transferring into a CSMP class. The largest difference was in the Comparison Test I, late-enterers scoring about a point below the other students.

The CSMP tests present a different picture. There are much larger differences; late-enterers scored at least two and as much as four points lower at all ability levels on Tests C and D and at the lower ability level for Test B. While the differences existed on almost all the individual subtests, they were largest with those dealing with integer calculations (Subtest 4), within certain arrow diagram problems (Subtests 10, 11 and 15), and with choosing the larger answer given two fractional computations (parts of Subtest 3). The items of these various subtests are shown in Tables 33, 39, 40, 42 and 32 respectively.

As previously noted, because of the limited number of students who could be tested individually, only students who had been in CSMP since first grade were tested with the Minicomputer. Hence it is not known whether late-enterers caught up with other students in this area. It should also be noted that the data provide no answers whatsoever about the relative progress of students where an entire class begins to use the CSMP class for the first time in the second grade.

Performance By Ability Levels

All CSMP students who had a mean pretest score below 40 were combined into one group and mean pretest and test scores computed for this group. This was also done for students with a mean pretest score of over 60. The same thing was done for Non-CSMP students. One can then separately compare CSMP and Non-CSMP performance for low ability and then for high ability students. For example, CSMP students had higher scores on the standardized tests; were they higher at only one level of ability or "across the board". The data are presented in Table 48.

Table 48

Test Scores by Ability Groups*

Test		Percentage of Students		Mean Pretest Score		Mean Test Score	
		Low	High	Low	High	Low	High
Comparison Test I	CSMP	23	15	31.3	67.7	14.6	27.9
	Non-CSMP	23	23	29.5	65.9	12.7	21.2
Comparison Test II	CSMP	16	30	28.2	66.6	21.4	45.9
	Non-CSMP	10	21	34.2	67.7	28.3	43.6
Comparison Test III	CSMP	14	27	33.0	67.2	22.7	29.6
	Non-CSMP	9	27	36.6	67.5	24.0	27.6
Comparison Test IV	CSMP	8	29	34.9	66.2	12.1	26.7
	Non-CSMP	5	37	34.0	67.5	10.8	24.5
CTBS: Computation	CSMP	19	24	30.1	66.5	14.5	25.6
	Non-CSMP	16	24	30.3	66.8	13.5	25.1
CTBS: Concepts & Applications	CSMP	18	24	30.0	66.6	13.1	22.9
	Non-CSMP	16	24	30.4	66.8	11.9	22.4

*Low Ability Group: Students with pretest score <40
High Ability Group: Students with pretest score >60

The answer to the question raised above is that for the standardized tests, both high and low ability CSMP students scored higher than the appropriate Non-CSMP students, though the difference was larger for low ability students. Generally at both levels of ability, CSMP students outperformed the Non-CSMP students. The exceptions are for low ability students on Comparison Tests II and III where the Non-CSMP students did better, though in those cases the low group for CSMP was distinctly lower than for Non-CSMP as shown by the fairly large differences in pretest scores.

Summary and Discussion

Thirteen second grade classes in four districts in the local area studied the CSMP curriculum during the 1974-75 school year. A wide range of abilities and socio-economic status was represented by these classes (see Tables 1 and 5). For each CSMP class there was a comparison class, in the same school where possible and in an adjacent school where the CSMP class was the only second grade class. These classes were usually the same groups of students who were tested the year before in first grade in the CSMP - Non-CSMP comparisons described in a previous report.* In the fall a test of mental abilities was administered to all students and the scores from this test were used as covariates in the subsequent analysis of test data in order to adjust for possible differences in entering ability between CSMP and Non-CSMP classes. Because of the rather careful pairing of classes such differences were generally quite small.

Two kinds of tests, a standardized test and a series of comparison tests, were administered to CSMP and Non-CSMP classes according to a testing plan (Table 6) which ensured that representative pairs of classes would write each test. In the case of the standardized test all 13 pairs of classes were used. The standardized test used was the Mathematics Test, with two subtests, of the Comprehensive Test of Basic Skills. Four Comparison Tests with varying numbers of subtests were developed, two group-administered to the whole class and two administered to about seven representative students of each selected class. These comparison tasks were essentially problems posed in situations which were novel to both CSMP and Non-CSMP students. They were intended to be situations in which the techniques and ways of thinking about mathematics which are stressed in the CSMP curriculum might be used with more success than would normally be the case.

Table 49 summarizes the results of these tests. Mean scores across classes are given for the CSMP and for the Non-CSMP classes and those tests on which the differences were significant at the .05 level are indicated. A two-way Analysis of Covariance was used, with classes as the unit of analysis, and the resulting F-test had between 1 and 4, and 1 and 11 degrees of freedom.

*Evaluation Report 1-A-3: Final Summary Report Year 1

Table 49

Summary Data for Tests
Administered to CSMP and Non-CSMP Classes

Test	Number of Pairs of Classes	Mean for CSMP Classes	Mean for Non-CSMP Classes	Significant at .05 level
Comparison Test I				
a) Word Problems		4.79	4.06	
b) Number Puzzles		2.66	2.51	
c) Estimation: Calculation	6	5.18	5.03	
d) Estimation: Largest Number		5.26	4.44	✓
e) Showing Fractions		1.15	0.92	
Total		19.06	17.01	
Comparison Test II				
a) Equations: Construction		2.72	2.03	✓
b) Equations: Fluency	7	13.63	11.19	✓
c) Combinatorics		16.94	17.08	
d) Regrouping		6.34	6.55	
Total		39.62	38.86	✓
Comparison Test III				
a) Classification	12	17.98	17.23	
b) Binary Relations		8.05	8.12	
Total		26.03	25.36	
Comparison Test IV				
a) Number Patterns	11	10.11	9.28	✓
b) Functions		6.48	5.75	
c) Probability		4.59	3.99	✓
Total		21.18	19.01	✓
Comprehensive Test of Basic Skills				
a) Computation	13	21.63	20.28	✓
b) Concepts and Applications		18.96	18.12	✓
Total		40.61	38.47	✓

All differences which were significant were in favor of CSMP. These included both the Computation and Concepts parts of the standardized tests and several of the various Comparison Tests. With the exception of Comparison Test IVc), Probability, the four subtests on which CSMP classes did significantly better than Non-CSMP classes were what might be called "directly numerical" situations; numbers (or numerals) themselves were the stimuli. Other subtests in which the given stimuli were situations in which numbers (or more widely, mathematical thinking) were to be applied (Ia, Ie, IIc, IId, IIIa, IIIb) did not produce significant differences.

These results are very encouraging. CSMP students did better on both parts of the standardized test and on several novel tasks of a numerical nature. It is also important to note that the results were consistent across school districts (see Table 46) and, perhaps more importantly, across ability levels. It is not

the case that differences were due to gains by high ability students; low ability students also did better (see Table 47). In addition, students who transferred into CSMP at the beginning of second grade appeared to do as well as students who were in CSMP from first grade except on Comparison Test I where they had slightly lower scores (Table 48).

In the author's opinion there are two explanations worth considering other than the obvious one that the CSMP curriculum is responsible for these differences. The first explanation is that teachers were not randomly assigned to CSMP and Non-CSMP (indeed this is virtually impossible to accomplish in the pilot stages of any program implementation). It is therefore possible that CSMP teachers may be, as a group, more capable of producing student achievement. This is a plausible explanation and cannot be entirely refuted. However, in discussions with principals who had both a CSMP and a Non-CSMP teacher in their school, there was no support for such differences; indeed in one case the same teacher taught both classes and in another the principal thought the Non-CSMP teacher was a better teacher. It is also true that this selection factor would be stronger in the first year of this study when the very first teachers to use CSMP are chosen. The next year, the second grade teachers more or less inherit the program and the classes.

The other explanation may be that CSMP classes spend more time in math instruction than do Non-CSMP classes. Based on interview data*, the mean number of minutes per day on math instruction in the 13 CSMP classes was 50, which is slightly higher than usual. However, there was no relationship between number of minutes and achievement either with or without adjusting for differences in entering ability. This information is not available for Non-CSMP classes though it is known that, in at least some classes, school scheduling dictated that the same time be spent for all math instructions.

Thus, while these two alternatives cannot be ruled out, there are good reasons for discounting them as explanations for higher achievement by CSMP students. The so-called "novelty effect" is also discounted for two reasons. First, described in the report dealing with student interviews**, students were usually not aware that their math program was particularly different from what other students used. Second, data collected over the past two years from first grade students indicates that, as judged by student performance, teachers do as well if not better in their second year of teaching CSMP as they did the first year.***

One is led to believe from what is presented above that CSMP students did significantly better than Non-CSMP students on several measures and that this can be attributed to the CSMP program and not to other factors. The reader should bear in mind that these results can not necessarily be generalized to school systems or classes or teachers beyond those participating in this study (ie. "local classes"). That is one of the penalties for non-random (in this case voluntary) sampling of school districts, and within districts, teachers.

---

*Evaluation Report 2-C-3: Teacher Interviews, First Grade
**Evaluation Report 2-B-3: Student Interviews
***Evaluation Report 2-B-2: Readministration of First Grade Test Items

A series of tests dealing with the unique aspects of CSMP, content not already covered in one of the above tests, was also administered to CSMP classes. The results are more difficult to interpret because of the lack of standards, either through what "usually happens" (control groups) or what "should happen" (a priori standards of achievement laid down by the program). Thus the interpretation of the data here is the author's own subjective evaluation of what one should expect second grade CSMP students to be able to do. However, this evaluation is based on knowledge of curriculum, the contents of the various lessons and relative emphasis of topics; and on discussions with teachers, observations of classes and interviews with students.

Rather than reviewing and commenting on the results of each subtest, which has already been done, the discussion will center on general topics.

a) Multiplication. The level of achievement with simple, whole-number, single-digit, multiplication is quite satisfactory, as indicated also by the standardized test data. It is also very satisfactory in two areas not typically given much consideration in second grade. A surprising number of students were able to double or triple relatively large numbers; nearly half the students could calculate "2x37". And in calculating fractional parts of a whole number, specifically 1/nth of a number, students also did very well; about half the students could calculate "$\frac{1}{2}$x48" and "$\frac{1}{4}$x20."

b) Integers. Negative integers are not studied in other second grade curricula. Hence there is no basis of comparison and it may safely be said that whatever progress CSMP students make is a "gain" over Non-CSMP students. The level of success is judged to be adequate; 60% of the students could select the one true number sentence of four very similar and possibly confusing sentences involving negative integers (Table 33), though only about half the students could compute "$\hat{4}+\hat{7}$" and "$\hat{5}+\hat{4}$." However, the improvement from first grade is rather disappointing. There was only a moderate increase from last year in the percentage correct on those items repeated from last year's test.

c) Arrow Diagrams. The method of looking at percent correct is not entirely satisfactory for this topic. Generally these percentages are satisfactory. On the one hand there are surprisingly high percentages of students getting difficult items correct; for example the "Detective Story" items in Table 39 and the problems in Table 42 whose format was completely new to students. On the other hand there were many subtests on which up to 20% of the students had virtually no success. For example,
   i) 12% of the students got none of the four "sending valentine" problems in Table 44 and 15% got no more than one of the five relatively easy "I am more than you" arrows in Table 38.
   ii) For each of the three subtests requiring dots and arrows to be labelled (Table 40), over 20% of the students got none of the four test items correct.
   iii) In labelling dots in order to locate the dot for 83 in Table 43, 15% of the students made five or more errors.

d) Minicomputer. Again there are a very wide range of scores; many could do virtually everything asked and many others could do virtually nothing. Over 30% of the students made at most one error on this rather comprehensive, individually administered test and these students generally worked very rapidly. On the other hand when considering low ability students (the

bottom 20% of the CSMP students as measured by the pretest), only two of these 11 students could consistently read and write numbers on the Minicomputer, let alone even set up the Minicomputer for subtraction, multiplication or taking one-half of a number.

Students who transferred into a second grade CSMP class during the first two months of school do not, as a group, perform as well on these tests of CSMP content as students who were in the program from the beginning (Table 47). Thus while new students do not appear to be penalized as far as the standard skills and concepts, they do not, during the course of the year, "catch up" with the other students in acquiring those concepts specific to CSMP. This is not a surprising result.

In summary then, CSMP classes did better than Non-CSMP classes on both standardized measures and some "content-free" comparison tasks and this is true for both high and low ability students and for both original and transfer students. Although overall performance was generally adequate and at times praiseworthy for unique CSMP content, on those subtests involving arrow diagrams and the Minicomputer a considerable number of students had little or no success, and students who transferred into CSMP classes near the beginning of the year were not able to catch up with their classmates on CSMP concepts.

Appendix

Analysis of Covariance Tables

For each of the tests administered to both CSMP and Non-CSMP classes an analysis of covariance table, in rather standard form, is presented. To illustrate this format the first part of Table A·1 is reproduced below, dealing with the addition items from the Computation Test of The Comprehensive Tests of Basic Skills.

Test	Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F-Ratio
Addition Items	Error(ignoring regression)	12	1.59		
	Regression	1	.58	.58	6.33*
	Error(after regression)	11	1.01	.09	
	Schools	12	4.02	.34	3.65
	CSMP - Non-CSMP	1	3.31	3.31	36.09*

1. The first source of variation is the error, or residual, unexplained variance not accounted for by the additive analysis-of-variance model wherein class score is predicted by a general mean plus an effect due to school (the school from which the CSMP - Non-CSMP pair was drawn) plus an effect due to curriculum (CSMP versus Non-CSMP). This model does not take into account regression on the covariate (pretest).
2. The next term "Regression" shows the reduction in this error variance when the model is expanded to take into account differences in the pretest scores of the various classes. A large reduction relative to the original error variance indicates that the expanded model provides a much better interpretation of the data. In the above example, approximately one-third (.58 of 1.59) of the previously unexplained error variance can in fact be explained by taking differences in pretest scores into account.
3. The third term is the reduced error variance (the original error minus regression) and is the standard against which other variances are judged for significance.
4. The fourth term is the variation in class scores which can be explained by systematic differences from school to school after fitting the expanded model. This is not of particular interest in this study, except to note that this factor often did indeed account for a considerable portion of the variance and was worth including in the model.

5. The fifth term is the variation in class scores which can be explained by differences in curriculum (i.e. between CSMP and Non-CSMP) after fitting the expanded model.

Each Sum of Squares is divided by its number of degrees of freedom to obtain the Mean Sum of Squares, thus providing a sort of average of the amount of variation due to each source. The F-ratio for any particular source of variation is simply its Mean Sum of Squares divided by the Mean Sum of Squares due to unexplained error. If this ratio is large, i.e. if the systematic error due to that source of variation is much larger than due to random, unexplained error, then the result is significant. In the above example, the F-ratio for CSMP - Non-CSMP differences is $3.31 \div .09 = 36.09$ with 1 and 11 degrees of freedom. This is very large, the largest encountered in all the comparisons, and the probability that the differences were merely random errors is extremely small; i.e. the differences are significant. All F-ratios which are significant are indicated by an asterisk and any such significant differences in the CSMP - Non-CSMP comparisons are in favor of CSMP classes.

Table A.1

Analysis of Covariance:
Comprehensive Test of Basic Skills

Test	Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F-Ratio
Addition Items	Error(ignoring regression)	12	1.59		
	Regression	1	.58	.58	6.33*
	Error(after regression)	11	1.01	.09	
	Schools	12	4.02	.34	3.65
	CSMP - Non-CSMP	1	3.31	3.31	36.09*
Subtraction Items	Error(ignoring regression)	12	3.70		
	Regression	1	.52	.52	1.78
	Error(after regression)	11	3.18	.29	
	Schools	12	13.08	1.09	3.77
	CSMP - Non-CSMP	1	1.45	1.45	5.04*
Multiplication Items	Error(ignoring regression)	12	2.45		
	Regression	1	.05	.05	.23
	Error(after regression)	11	2.40	.22	
	Schools	12	6.57	.55	2.52
	CSMP - Non-CSMP	1	.28	.28	1.27
Total Computation	Error(ignoring regression	12	9.62		
	Regression	1	2.83	2.83	4.58
	Error(after regression)	11	6.79	.62	
	Schools	12	41.04	3.42	5.54*
	CSMP - Non-CSMP	1	11.68	11.68	18.92*
Concepts & Applications	Error(ignoring regression)	12	26.03		
	Regression	1	19.14	19.14	30.57*
	Error(after regression)	11	6.89	.63	
	Schools	12	26.45	2.20	3.52*
	CSMP - Non-CSMP	1	3.93	3.93	6.27*
Total Mathematics Test	Error(ignoring regression)	12	58.85		
	Regression	1	33.66	33.66	14.69*
	Error(after regression)	11	25.19	2.29	
	Schools	12	104.73	8.73	3.81*
	CSMP - Non-CSMP	1	32.64	32.64	14.25*

Table A·2

Analysis of Covariance
Comparison Test I

Test	Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F-Ratio
Ia) Word Problems	Error(ignoring regression)	5	3.47		
	Regression	1	1.09	1.09	1.85
	Error(after regression)	4	2.37	.59	
	Schools	5	3.21	.64	1.08
	CSMP - Non-CSMP	1	1.18	1.18	1.98
Ib) Number Puzzles	Error(ignoring regression)	5	.56		
	Regression	1	.12	.12	1.07
	Error(after regression)	4	.44	.11	
	Schools	5	.21	.04	.37
	CSMP - Non-CSMP	1	.04	.04	.39
Ic) Estimation-Calculation	Error(ignoring regression)	5	2.19		
	Regression	1	.66	.66	1.75
	Error(after regression)	4	1.54	.38	
	Schools	5	2.30	.46	1.21
	CSMP - Non-CSMP	1	.00	.00	.01
Id) Estimation-Largest Number	Error(ignoring regression	5	.94		
	Regression	1	.00	.00	.01
	Error(after regression)	4	.94	.23	
	Schools	5	.74	.15	.63
	CSMP - Non-CSMP	1	2.00	2.00	8.54*
Ie) Showing Fractions	Error(ignoring regression)	5	.26		
	Regression	1	.06	.06	1.10
	Error(after regression)	4	.20	.05	
	Schools	5	.06	.01	.24
	CSMP - Non-CSMP	1	.14	.14	2.67
I Total	Error(ignoring regression)	5	16.01		
	Regression	1	5.55	5.55	2.12
	Error(after regression)	4	10.46	2.62	
	Schools	5	13.06	2.61	1.00
	CSMP - Non-CSMP	1	10.02	10.02	3.83

Table A•3

Analysis of Covariance
Comparison Test II

Test	Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F-Ratio
IIa) Equations-Construction	Error(ignoring regression)	6	.36		
	Regression	1	.02	.02	.33
	Error(after regression)	5	.33	.07	
	Schools	6	.11	.02	.28*
	CSMP - Non-CSMP	1	1.63	1.63	24.44*
IIb) Equations-Fluency	Error(ignoring regression)	6	9.06		
	Regression	1	2.29	2.29	1.69
	Error(after regression)	5	6.77	1.35	
	Schools	6	10.13	1.69	1.25
	CSMP - Non-CSMP	1	19.18	19.18	14.16*
IIc) Combinatorics	Error(ignoring regression)	6	6.43		
	Regression	1	.99	.99	.91
	Error(after regression)	5	5.44	1.09	
	Schools	6	.77	.13	.12
	CSMP - Non-CSMP	1	.14	.14	.13
IId) Reasoning	Error(ignoring regression)	6	7.40		
	Regression	1	5.29	5.29	10.98*
	Error(after regression)	5	2.12	.46	
	Schools	6	4.39	.73	1.58
	CSMP - Non-CSMP	1	.39	.39	.85
II Total	Error(ignoring regression	6	37.53		
	Regression	1	23.60	23.60	8.47*
	Error(after regression)	5	13.93	2.79	
	Schools	6	11.39	1.90	.68
	CSMP - Non-CSMP	1	21.09	21.09	7.57*

Table A.4

Analysis of Covariance
Comparison Test III

Test	Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F-Ratio
IIIa) Classification	Error(ignoring regression	11	16.46		
	Regression	1	.90	.90	.58
	Error(after regression)	10	15.56	1.56	
	Schools	11	35.50	3.23	2.07
	CSMP - Non-CSMP	1	1.34	1.34	.86
IIIb) Binary Relations	Error(ignoring regression)	11	5.34		
	Regression	1	1.36	1.36	3.42
	Error(after regression)	10	3.98	.40	
	Schools	11	22.05	2.00	5.04*
	CSMP - Non-CSMP	1	.51	.51	1.29
III Total	Error(ignoring regression)	11	17.82		
	Regression	1	4.47	4.47	3.35
	Error(after regression)	10	13.34	1.33	
	Schools	11	95.27	8.66	6.49*
	CSMP - Non-CSMP	1	.19	.19	.14

Table A·5

Analysis of Covariance
Comparison Test IV

Test	Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F-Ratio
IVa) Number Patterns	Error(ignoring regression)	10	7.34		
	Regression	1	.27	.27	.34
	Error(after regression)	9	7.07	.79	
	Schools	10	19.24	1.92	2.45
	CSMP - Non-CSMP	1	3.99	3.99	5.08*
IVb) Functions	Error(ignoring regression)	10	10.83		
	Regression	1	1.71	1.71	1.69
	Error(after regression)	9	9.12	1.01	
	Schools	10	16.06	1.61	1.59
	CSMP - Non-CSMP	1	3.40	3.40	3.36
IVc) Probability	Error(ignoring regression)	10	3.86		
	Regression	1	1.11	1.11	3.64*
	Error(after regression)	9	2.75	.31	
	Schools	10	8.25	.82	2.70
	CSMP - Non-CSMP	1	1.65	1.65	5.40*
IV Total	Error(ignoring regression)	10	26.25		
	Regression	1	.60	.60	.21
	Error(after regression)	9	25.65	2.85	
	Schools	10	78.18	7.82	2.74
	CSMP - Non-CSMP	1	26.47	26.47	9.29*